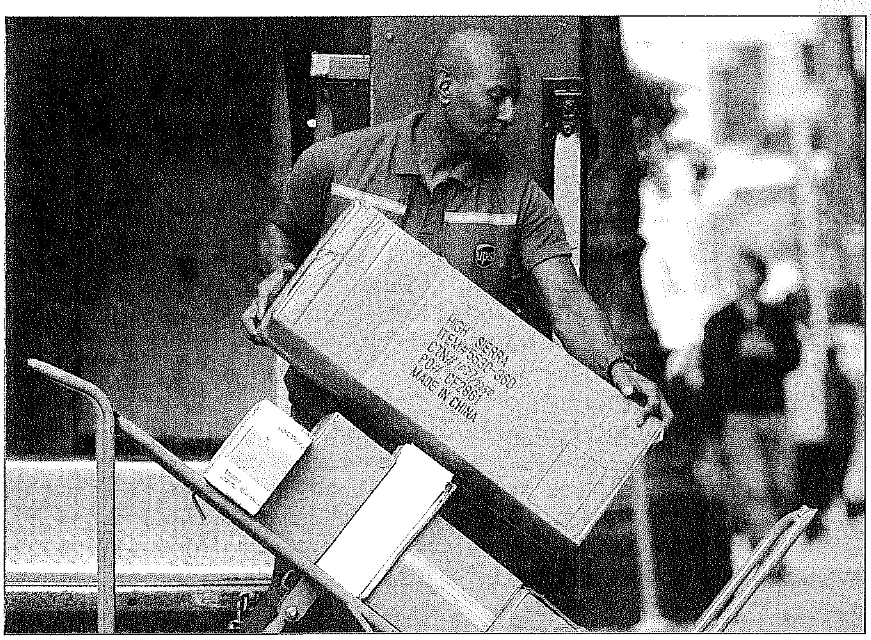
3

Descripción de datos

Medidas numéricas



Los pesos (en libras) de una muestra de cinco cajas que se envían por UPS son los siguientes: 12, 6, 7, 3 y 10. Calcule la desviación estándar (vea ejercicio 76 y objetivo 4).

OBJETIVOS

Al concluir el capítulo, será capaz de:

1. Calcular la *media aritmética*, la *media ponderada*, la *mediana*, la *moda* y la *media geométrica*.

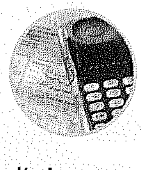
2. Explicar las características, usos, ventajas y desventajas de cada *medida de ubicación*.

3. Identificar la posición de la media, la mediana y la moda para las *distribuciones simétrica y sesgada*.

4. Calcular e interpretar el *rango*, la *desviación media*, la *varianza* y la *desviación estándar*.

5. Comprender las características, usos, ventajas y desventajas de cada *medida de dispersión*.

6. Comprender el teorema de *Chebyshev* y la *regla empírica* en relación con un conjunto de observaciones.



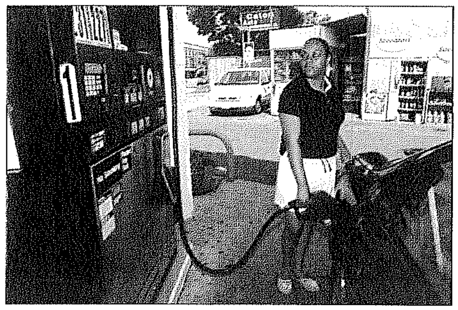
Estadística en acción

¿Se ha topado alguna vez con un estadounidense *promedio*? Pues bien, se llama Robert (nivel nominal de la medición); tiene 31 años (nivel de razón); mide 1.77 metros (otro nivel de razón de la medición); pesa 78 kilogramos; calza del 9½; su cintura mide 85 cm de diámetro y viste trajes talla 40. Además, el hombre promedio come 1.8 kg de papas fritas; mira 2 567 horas el televisor y se come 11.77 kg de plátanos al año, además de que duerme 7.7 horas cada noche.

La estadounidense promedio mide 1.64 metros de estatura y pesa 64 kg, mientras que la modelo estadounidense promedio mide 1.65 metros y pesa 53 kg. Un día cualquiera, casi la mitad de las mujeres en Estados Unidos está a dieta. Idolatrada en la década de los cincuenta, Marilyn Monroe se consideraría con sobrepeso según los estándares de hoy. Usaba vestidos de las tallas 14 a la 18, y era una mujer saludable y atractiva.

Introducción

El capítulo 2 inicia al estudio de la estadística descriptiva. Para transformar un cúmulo de datos en bruto en algo con significado, primero debe organizar los datos cuantitativos en una distribución de frecuencias y después hacer una representación gráfica como un histograma; hay otras técnicas para graficar, como las gráficas de pastel, útil para representar datos cualitativos, y polígonos de frecuencias para representar datos cuantitativos.



Este capítulo presenta dos formas numéricas de describir datos cuantitativos: las **medidas de ubicación** y las **medidas de dispersión.** A las medidas de ubicación a menudo se les llama **promedios.** El propósito de una medida de ubicación consiste en señalar el centro de un conjunto de valores.

Usted está familiarizado con el concepto de promedio, medida de ubicación que muestra el valor central de los datos. Los promedios aparecen diario en televisión, en el periódico y otras publicaciones. He aquí algunos ejemplos:

- La casa promedio en Estados Unidos cambia de dueño cada 11.8 años.
- El precio promedio de un galón de gasolina, la semana pasada, en Carolina del Sur era de $2.47 de acuerdo con un estudio de la Asociación Estadounidense de Automóviles.
- El costo promedio por conducir un automóvil particular es de $10 361 anuales en Los Ángeles; de $9 660 anuales en Boston; de $10 762 anuales en Filadelfia.
- Un estadounidense recibe un promedio de 568 piezas de correspondencia cada año.
- El salario inicial promedio para un graduado de la escuela de administración el año pasado era de $38 254. Para un graduado con licenciatura en artes liberales, era de $30 212.
- Hay 26.4 millones de golfistas mayores de 12 años en Estados Unidos. Cerca de 6.1 millones son fervientes golfistas; es decir que juegan un promedio de 25 partidos al año. Más información relacionada con los golfistas: el costo medio de un partido de golf en un campo público de 18 hoyos en Estados Unidos es de $30. Hoy día, el típico golfista es hombre, de 40 años de edad, con un ingreso familiar de $68 209.
- En Chicago la temperatura media alta es de 84 grados en julio y de 31 grados en enero. La precipitación media es de 3.80 pulgadas en julio y de 1.90 pulgadas en enero.

Si sólo toma en cuenta las medidas de ubicación en un conjunto de datos o si compara varios conjuntos de datos utilizando valores centrales, llegará a una conclusión incorrecta. Además de las medidas de ubicación, debe tomar en consideración la **dispersión**, denominada con frecuencia *variación* o *propagación*, en los datos. Por ejemplo, suponga que el ingreso anual promedio de los ejecutivos de compañías relacionadas con Internet es de $80 000 y que el ingreso promedio de ejecutivos de compañías farmacéuticas es también de $80 000. Si sólo atiende a los ingresos promedio, podría concluir, equivocadamente, que las dos distribuciones de salarios son idénticas o casi idénticas. Un vistazo a los rangos salariales indica que esta conclusión no es correcta. Los salarios de los ejecutivos en las empresas de Internet van de $70 000 a $90 000, en cambio los salarios de los ejecutivos de marketing de la industria farmacéutica van de $40 000 a $120 000. Por consiguiente, aunque los salarios promedio son los mismos en las dos industrias, hay más propagación o dispersión en los salarios de los ejecutivos de la industria farmacéutica. Para describir la dispersión considere el rango, la desviación media, la varianza y la desviación estándar.

En principio se discuten las medidas de ubicación. No existe una medida de dispersión; de hecho, existen varias. Consideraremos cinco: la media aritmética, la media ponderada, la mediana, la moda y la media geométrica. La media aritmética es la medida de ubicación que más se utiliza y que se publica con mayor frecuencia. Considerará la media como parámetro de población y como estadístico de las muestras.

La media poblacional

Muchos estudios incluyen todos los valores que hay en una población. Por ejemplo, hay 39 salidas en la carretera interestatal 75, que pasa por el estado de Kentucky. La distancia media entre dichas salidas es de 4.76 millas. Éste es el parámetro poblacional, ya que es la distancia entre *todas* las salidas. Hay 12 asociados de ventas empleados en la tienda de menudeo Reynolds Road, de Carpets by Otto. El monto promedio de comisiones que ganaron el mes pasado fue de $1 345. Éste es el valor poblacional, puesto que considera la comisión de *todos* los asociados de ventas. Otros ejemplos de media poblacional serían los siguientes: el precio de cierre promedio de las acciones de Johnson & Johnson durante los últimos 5 días es de $61.75; la tasa anual promedio de recuperación durante los últimos 10 años de Berger Funds es de 8.67% y el promedio de horas extra que trabajaron la semana pasada los seis soldadores del departamento de soldadura de Butts Welding, Inc., es de 6.45 horas.

En el caso de los datos en bruto, que no han sido agrupados en una distribución de frecuencias, la media poblacional es la suma de todos los valores en la población dividida entre el número de valores de la población. Para determinar la media poblacional, aplique la siguiente fórmula:

$$\text{Media poblacional} = \frac{\text{Suma de todos los valores en la población}}{\text{Número de valores en la población}}$$

En lugar de escribir las instrucciones completas para calcular la media poblacional (o cualquier otra medida), resulta más conveniente utilizar símbolos matemáticos adecuados. La media de una población con símbolos matemáticos es

MEDIA POBLACIONAL $$\mu = \frac{\Sigma X}{N} \qquad \textbf{[3.1]}$$

en la cual:

- μ representa la media poblacional; se trata de la letra minúscula griega *mu;*
- N es el número de valores en la población;
- X representa cualquier valor particular;
- Σ es la letra mayúscula griega *sigma* e indica la operación de suma;
- ΣX es la suma de X valores en la población.

Cualquier característica medible de una población recibe el nombre de **parámetro**. La media de una población es un parámetro.

PARÁMETRO Característica de una población.

Ejemplo

Hay 12 compañías fabricantes de automóviles en Estados Unidos. Enseguida aparece la lista del número de patentes concedidas por el Gobierno de Estados Unidos a cada compañía en un año reciente.

Compañía	Número de patentes concedidas	Compañía	Número de patentes concedidas
General Motors	511	Mazda	210
Nissan	385	Chrysler	97
DaimlerChrysler	275	Porsche	50
Toyota	257	Mitsubishi	36
Honda	249	Volvo	23
Ford	234	BMW	13

¿Representa esta información una muestra o una población? ¿Cuál es la media aritmética del número de patentes concedidas?

Solución

Es una población, ya que se toma en cuenta a todas las compañías fabricantes que consiguen patentes. Sume el número de patentes de cada una de las 12 compañías. El número total de patentes de las 12 compañías es de 2 340. Para determinar la media aritmética, divida este total entre 12. Así, la media aritmética es 195, calculada mediante la operación 2 340/12. De acuerdo con la fórmula 3.1,

$$\mu = \frac{511 + 385 + \cdots + 13}{12} = \frac{2\,340}{12} = 195$$

¿Cómo interpretar el valor 195? El número típico de patentes que recibe una compañía fabricante de automóviles es 195. Como se ha tomado en cuenta a todas las compañías que reciben patentes, este valor es un parámetro poblacional.

Media de una muestra

Como se explicó en el capítulo 1, con frecuencia se selecciona una muestra de la población para encontrar algo sobre una característica específica de la población. Por ejemplo, el departamento de control de calidad necesita asegurarse de que los rodamientos de balas fabricados tengan un diámetro externo aceptable. Resultaría muy costoso y consumiría demasiado tiempo verificar el diámetro externo de todos los rodamientos producidos. Por consiguiente, se selecciona una muestra de cinco rodamientos y se calcula el diámetro externo de cinco rodamientos para aproximar el diámetro medio de todos.

En el caso de los datos en bruto, de los datos no agrupados, *la media es la suma de los valores de la muestra, divididos entre el número total de valores de la muestra*. La media de una muestra se determina de la siguiente manera:

Media de datos no agrupados de una muestra

$$\text{Media de la muestra} = \frac{\text{Suma de todos los valores de la muestra}}{\text{Número de valores de la muestra}}$$

La media muestral y la media poblacional se calculan en la misma manera, pero la notación abreviada que se emplea es diferente. La fórmula de la media muestral es:

MEDIA DE UNA MUESTRA $$\bar{X} = \frac{\Sigma X}{n} \qquad [3.2]$$

en la cual:

$\bar{X}$ es la media de la muestra; se lee: X barra;
n es el número de valores de la muestra.

La media de una muestra o cualquier otra medición basada en una muestra de datos recibe el nombre de **estadístico.** Si el diámetro promedio externo de una muestra de cinco rodamientos de bala es de 0.625 pulgadas, se trata de un ejemplo de estadístico.

ESTADÍSTICO Característica de una muestra.

Ejemplo

SunCom estudia la cantidad de minutos que emplean sus clientes en un plan tarifario de cierto teléfono celular. Una muestra aleatoria de 12 clientes arroja la siguiente cantidad de minutos empleados el mes pasado.

90	77	94	89	119	112
91	110	92	100	113	83

¿Cuál es valor de la media aritmética de los minutos empleados?

Solución

De acuerdo con la fórmula 3.2, la media muestral es:

$$\text{Media muestral} = \frac{\text{Suma de todos los valores en la muestra}}{\text{Número de valores en la muestra}}$$

$$\bar{X} = \frac{\Sigma X}{n} = \frac{90 + 77 + \cdots + 83}{12} = \frac{1\,170}{12} = 97.5$$

El valor de la media aritmética de los minutos empleados el mes pasado por los usuarios de teléfonos celulares de la muestra es de 97.5 minutos.

Propiedades de la media aritmética

La media aritmética es una medida de ubicación muy utilizada. Cuenta con algunas propiedades importantes:

1. **Todo conjunto de datos de intervalo –o de nivel de razón– posee una media**. Recuerde del capítulo 1 que los datos del nivel de razón incluyen datos como edades, ingresos y pesos, en éstos la distancia entre los números es constante.
2. **Todos los valores se encuentran incluidos en el cálculo de la media**.
3. **La media es única**. Sólo existe una media en un conjunto de datos. Más adelante en el capítulo descubrirá un promedio que podría aparecer dos o más veces en un conjunto de datos.
4. **La suma de las desviaciones de cada valor de la media es cero**. Expresado simbólicamente,

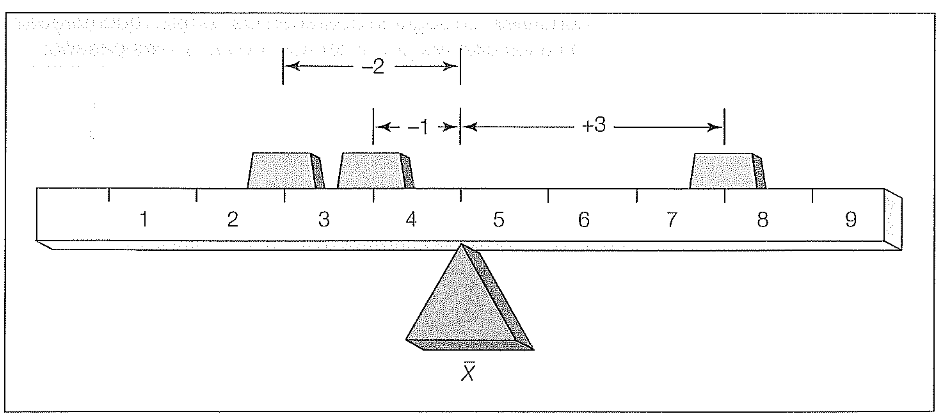
$$\Sigma(X - \bar{X}) = 0$$

Como ejemplo, la media de 3, 8 y 4 es 5. De esta manera,

$$\begin{aligned}\Sigma(X - \bar{X}) &= (3 - 5) + (8 - 5) + (4 - 5)\\ &= -2 + 3 - 1\\ &= 0\end{aligned}$$

La media como punto de equilibro

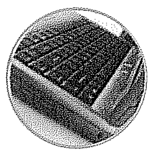
De esta manera la media es un punto de equilibrio de un conjunto de datos. Para ilustrarlo, imagine una regla con los números 1, 2, 3, ...,9 uniformemente espaciados. Suponga que se colocaran tres barras del mismo peso sobre la regla en los números 3, 4 y 8 y que el punto de equilibrio se colocara en 5, la media de los tres números. Descubriría que la regla se equilibra perfectamente. Las desviaciones debajo de la media (–3) son iguales a las desviaciones por encima de la media (+3). El esquema es:



La media se ve afectada en exceso por valores grandes o pequeños poco comunes

La media tiene un punto débil. Recuerde que el valor de cada elemento en una muestra, o población, se utiliza cuando se calcula la media. Si uno o dos de estos valores son extremadamente grandes o pequeños comparados con la mayoría de los datos, la media podría no ser un promedio adecuado para representar los datos. Por ejemplo, suponga que los ingresos anuales de un pequeño grupo de corredores de bolsa en Merrill Lynch es de $62 900, $61 600, $62 500, $60 800 y $1 200 000. El ingreso medio es de $289 560; claro, no es representativo del grupo, ya que todos, salvo un corredor, tienen ingresos entre $60 000 y $63 000. Un ingreso ($1.2 millones) afecta en exceso la media.

Autoevaluación 3.1



1. Los ingresos anuales de una muestra de empleados de gerencia media en Westinghouse son: $62 900, $69 100, $58 300 y $76 800.
   a) Proporcione una fórmula para la media muestral.
   b) Determine la media muestral.
   c) ¿Es la media que calculó en el inciso *b*) un estadístico o un parámetro? ¿Por qué razón?
   d) ¿Cuál es su mejor aproximación de la media de la población?
2. Todos los estudiantes de Ciencias Avanzadas de la Computación de la clase 411 constituyen una población. Sus calificaciones en el curso son de 92, 96, 61, 86, 79 y 84.
   a) Proporcione la fórmula de la media poblacional.
   b) Calcule la calificación media del curso.
   c) ¿Es la media que calculó en el inciso *b*) un estadístico o un parámetro? ¿Por qué razón?

Ejercicios

Las respuestas a los ejercicios impares se encuentran al final del libro.

1. Calcule la media de la siguiente población de valores: 6, 3, 5, 7, 6.
2. Calcule la media de la siguiente población de valores: 7, 5, 7, 3, 7, 4.
3. ***a*)** Calcule la media de los siguientes valores muestrales: 5, 9, 4, 10.
***b*)** Demuestre que $\Sigma(X - X) = 0$.
4. ***a*)** Calcule la media de los siguientes valores muestrales: 1.3, 7.0, 3.6, 4.1, 5.0.
***b*)** Demuestre que $\Sigma(X - X) = 0$.
5. Calcule la media de los siguientes valores muestrales: 16.25, 12.91, 14.58.
6. Calcule el salario promedio por hora pagado a carpinteros que ganan los siguientes salarios por hora: $15.40, $20.10, $18.75, $22.76, $30.67, $18.00.

En los ejercicios 7 a 10, *a*) calcule la media aritmética y *b*) indique si se trata de un estadístico o de un parámetro.

7. Midtown Ford emplea a 10 vendedores. El número de automóviles nuevos vendidos el mes pasado por los respectivos vendedores fueron: 15, 23, 4, 19, 18, 10, 10, 8, 28, 19.
8. El departamento de contabilidad en una compañía de ventas por catálogo contó las siguientes cantidades de llamadas recibidas por día en el número gratuito de la compañía durante los primeros 7 días de mayo de 2006: 14, 24, 19, 31, 36, 26, 17.
9. Cambridge Power and Light Company seleccionó una muestra aleatoria de 20 clientes residenciales. En seguida aparecen las sumas, redondeadas al dólar más próximo, que se cobraron a los clientes por el servicio de luz el mes pasado:

54	48	58	50	25	47	75	46	60	70
67	68	39	35	56	66	33	62	65	67

10. El director de relaciones humanas de Ford inició un estudio de las horas de trabajo extra en el Departamento de Inspección. Una muestra de 15 trabajadores reveló que éstos laboraron la siguiente cantidad de horas extra el mes pasado.

13	13	12	15	7	15	5	12
6	7	12	10	9	13	12	

11. AAA Heating and Air Conditioning concluyó 30 trabajos el mes pasado con un ingreso medio de $5 430 por trabajo. El presidente desea conocer el ingreso total del mes. Sobre la base de la información limitada, ¿puede calcular el ingreso total? ¿A cuánto asciende?

12. Una compañía farmacéutica grande contrata graduados de administración de empresas para vender sus productos. La compañía se expande rápidamente y dedica un día a capacitar en ventas a los nuevos vendedores. El objetivo que la compañía fija a cada nuevo vendedor es de $10 000 mensuales. Éste se basa en las ventas promedio actuales de toda la compañía, que son de $10 000 mensuales. Después de revisar las retenciones de impuestos de los nuevos empleados, la compañía encuentra que sólo 1 de cada 10 empleados permanece más de tres meses en la empresa. Haga algún comentario sobre la utilización de las ventas promedio actuales mensuales como objetivo de ventas para los nuevos empleados. ¿Por qué abandonan los empleados la compañía?

Media ponderada

La media ponderada constituye un caso especial de la media aritmética y se presenta cuando hay varias observaciones con el mismo valor. Para explicar esto, suponga que el Wendy´s Restaurant vende refrescos medianos, grandes y gigantes a $0.90, $1.25 y $1.50. De las 10 últimas bebidas vendidas 3 eran medianas, 4 grandes y 3 gigantes. Para determinar el precio promedio de las últimas 10 bebidas vendidas recurra a la fórmula 3.2.

$$\bar{X} = \frac{\$.90 + \$.90 + \$.90 + \$1.25 + \$1.25 + \$1.25 + \$1.25 + \$1.50 + \$1.50 + \$1.50}{10}$$

$$\bar{X} = \frac{\$12.20}{10} = \$1.22$$

el precio promedio de venta de las últimas 10 bebidas es de $1. 22.

Una manera fácil para determinar el precio promedio de venta consiste en determinar la media ponderada; multiplique cada observación por el número de veces que aparece. La media ponderada se representa como $\bar{X}_w$, que se lee: "X subíndice w".

$$\bar{X}_w = \frac{3(\$0.90) + 4(\$1.25) + 3(\$1.50)}{10} = \frac{\$12.20}{10} = \$1.22$$

En este caso las ponderaciones son conteos de frecuencias. Sin embargo, cualquier medida de importancia podría utilizarse como una ponderación. En general, la media ponderada del conjunto de números representados como $X_1, X_2, X_3, \ldots, X_n$ con las ponderaciones correspondientes $w_1, w_2, w_3, \ldots, w_n$, se calcula de la siguiente manera:

MEDIA PONDERADA $$\bar{X}_w = \frac{w_1X_1 + w_2X_2 + w_3X_3 + \cdots + w_nX_n}{w_1 + w_2 + w_2 + \cdots + w_n} \qquad \textbf{[3.3]}$$

La cual se abrevia de la siguiente manera:

$$\bar{X}_w = \frac{\Sigma(wX)}{\Sigma w}$$

Observe que el denominador de una media ponderada siempre es la suma de las ponderaciones.

Ejemplo

Carter Construction Company paga a sus empleados que trabajan por hora $16.50, $19.50 o $25.00 la hora. Hay 26 empleados contratados para trabajar por hora, 14 de los cuales reciben una paga con la tarifa de $16.50; 10 con la tarifa de $19.00 y 2 con la de $25.00. ¿Cuál es la tarifa promedio por hora que se paga a los 26 empleados?

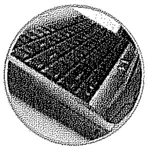
Solución

Para determinar la tarifa media por hora, multiplique cada una de las tarifas por hora por el número de empleados que ganan dicha tarifa. De acuerdo con la fórmula 3.3, la tarifa media por hora es:

$$\bar{X}_w = \frac{14(\$16.50) + 10(\$19.00) + 2(\$25.00)}{14 + 10 + 2} = \frac{\$471.00}{26} = \$18.1154$$

El salario promedio ponderado por hora se redondea a $18.12.

Autoevaluación 3.2



Springers vendió 95 trajes para caballero Antonelli a un precio normal de $400. Para la venta de primavera rebajaron los trajes a $200 y vendieron 126. Al final de la venta de liquidación, redujeron el precio a $100 y los restantes 79 trajes fueron vendidos.

a) ¿Cuál fue el precio promedio ponderado de un traje Antonelli?

b) Springers pagó $200 por cada uno de los 300 trajes. Haga algún comentario sobre la ganancia de la tienda por traje, si un vendedor recibe $25 de comisión por cada traje que vende.

Ejercicios

13. En junio una inversionista compró 300 acciones de Oracle (una compañía de tecnología de la información) a $20 la acción. En agosto compró 400 acciones más a $25 cada una. En noviembre compró otras 400 acciones, pero el precio bajó a $23 la acción. ¿Cuál es el precio promedio ponderado de cada acción?

14. Bookstall, Inc., es una librería especializada que se dedica a la venta de libros usados por Internet. Los libros de pasta blanda cuestan $1.00 cada uno y los de pasta dura, $3.50 cada uno. De los 50 libros vendidos el pasado martes por la mañana, 40 eran de pasta blanda y el resto de pasta dura. ¿Cuál fue el precio promedio ponderado de un libro?

15. Loris Healthcare System tiene 200 empleados en su personal de enfermería. Cincuenta son auxiliares de enfermería; 50 enfermeras practicantes y 100 son enfermeras tituladas. Las auxiliares de enfermería ganan $8 la hora; las enfermeras practicantes $15 la hora y las tituladas $24 la hora. ¿Cuál es el salario promedio ponderado por hora?

16. Andrews and Associates se especializa en leyes empresariales. Cobran $100 la hora de investigación de un caso; $75 la hora de asesoría y $200 la hora de redacción de un expediente. La semana pasada uno de los socios dedicó 10 horas a dar asesoría a una clienta, 10 horas a la investigación del caso y 20 horas a la redacción del expediente. ¿Cuál fue el monto medio ponderado por hora de honorarios por servicios legales?

Mediana

Ya se ha insistido en que si los datos contienen uno o dos valores muy grandes o muy pequeños, la media aritmética no resulta representativa. Es posible describir el centro de dichos datos a partir de una medida de ubicación denominada **mediana**.

Para ilustrar la necesidad de una medida de ubicación diferente de la media aritmética, suponga que busca un condominio en Palm Aire. Su agente de bienes raíces le dice que el precio típico de las unidades disponibles en este momento es de $110 000. ¿Aún insiste en seguir buscando? Si usted se ha fijado un presupuesto máximo de $75 000, podría pensar que los condominios se encuentran fuera de su presupuesto. Sin embargo, la verificación de los precios de las unidades individuales podría hacerle cambiar de parecer. Los costos son de $60 000, $65 000, $70 000, $80 000 y de $275 000 en el caso de un lujoso penthouse. El importe promedio aritmético es de $110 000, como le informó el agente de bienes raíces, pero un precio ($275 000) eleva la media aritmética y lo convierte en un promedio no representativo. Parece que un precio de poco más o menos $70 000 es un promedio más típico o representativo, y así es. En casos como éste, la mediana proporciona una medida de ubicación más válida.

MEDIANA Punto medio de los valores una vez que se han ordenado de menor a mayor o de mayor a menor.

El precio mediano de las unidades disponibles es de $70 000 Para determinarlo, ordene los precios de menor ($60 000) a mayor ($275 000) y seleccione el valor medio

($70 000). En el caso de la mediana los datos deben ser por lo menos de un nivel ordinal de medición.

Precios ordenados de menor a mayor		Precios ordenados de mayor a menor
$ 60 000		$275 000
65 000		80 000
70 000	← Mediana →	70 000
80 000		65 000
275 000		60 000

A la mediana le afectan menos los valores extremos

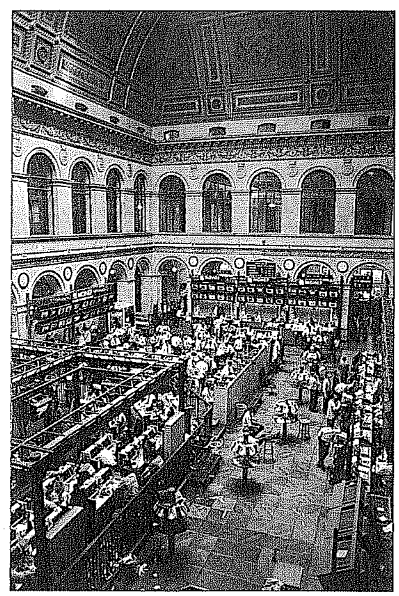
Observe que existe el mismo número de precios bajo la mediana de $70 000 que sobre ella. Por consiguiente, a la mediana no le afectan precios bajos o altos. Si el precio más alto fuera de $90 000 o de $300 000, incluso de $1 000 000, el precio mediano aún sería de $70 000. Asimismo, si el precio más bajo fuera de $20 000 o $50 000, el precio mediano todavía sería de $70 000.

En el ejemplo anterior hay un número *impar* de observaciones (cinco). ¿Cómo se determina la mediana en el caso de un número *par* de observaciones? Como antes, se ordenan las observaciones. Enseguida, con el fin de obtener un único valor por convención, calcule la media de las dos observaciones medias. Así, en el caso de un número par de observaciones, la mediana quizá no sea uno de los valores dados.

Ejemplo

Los rendimientos totales de tres años de los mejores fondos mutualistas accionarios de más alto desempeño se enlistan en seguida. ¿Cuál es el rendimiento mediano anualizado?

Solución



Nombre del fondo	Rendimiento total anualizado
Artisian Mid Cap	42.10%
Clipper	15.50
Fidelity Advisor Mid-Cap	27.58
Fidelity Mid-Cap Stock	28.64
Smith Barney Aggressive	41.77
Van Kampen Comstock	16.97

Observe que el número de rendimientos es *par* (6). Como hizo antes, primero ordene los rendimientos de menor a mayor. Enseguida identifique los dos rendimientos de en medio. La media aritmética de las dos observaciones de en medio proporciona el rendimiento mediano. Ordenados del más bajo al más alto, quedan:

Clipper	15.50%	
Van Kampen Comstock	16.97	
Fidelity Advisor Mid-Cap	27.58	56.22/2 =
Fidelity Mid-Cap Stock	28.64	28.11 %
Smith Barney Aggressive	41.77	
Artisian Mid Cap	42.10	

Preste atención a que la mediana no es uno de los valores. Asimismo, la mitad de los rendimientos se encuentran por debajo de la mediana y la mitad sobre ella.

Las principales propiedades de la mediana son las siguientes:

La mediana se determina para cualquier nivel de datos, excepto los nominales

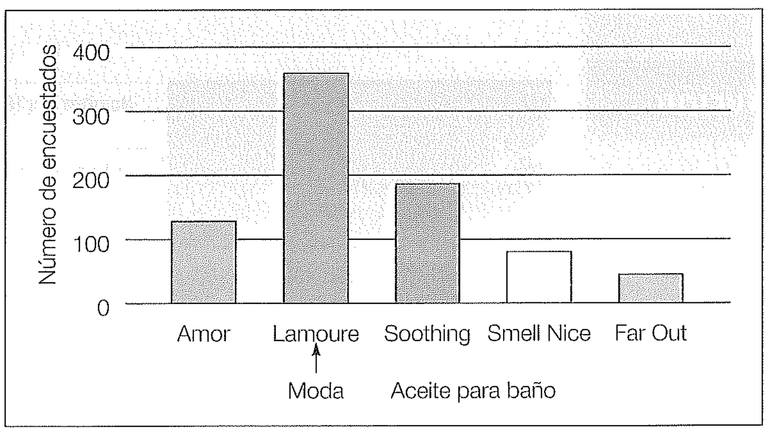
1. **No influyen en ella valores extremadamente grandes o pequeños.** Por consiguiente, la mediana es una valiosa medida de ubicación cuando dichos valores se presentan.
2. **Es calculable para datos de nivel ordinal o más altos.** Recuerde que en el capítulo 1 se ordenaron los datos de nivel ordinal de menor a mayor, como las respuestas *excelente, muy bien, bien, aceptable* y *mal* a una pregunta de una encuesta de mercado. Para dar un ejemplo sencillo, suponga que cinco personas califican una nueva barra de dulce de leche. Una persona pensó que era excelente; otra, muy buena; la siguiente la calificó de buena; una más, de aceptable y la quinta la consideró mala. La respuesta mediana es *buena.* La mitad de las respuestas se encuentran por encima de *buena;* la otra mitad por debajo.

Moda

La **moda** es otra medida de ubicación.

MODA Valor de la observación que aparece con mayor frecuencia.

La moda es de especial utilidad para resumir datos de nivel nominal. Un ejemplo de esta aplicación en datos de nivel nominal: una compañía creó cinco aceites para baño. La gráfica de barras 3.1 muestra los resultados de una encuesta de mercado diseñada para determinar qué aceite para baño prefieren los consumidores. La mayoría de los encuestados se inclinó por Lamoure, según lo evidencia la barra más grande. Por consiguiente, Lamoure representa la moda.



GRÁFICA 3.1 Número de encuestados que prefieren ciertos aceites para baño

Ejemplo

Los salarios anuales de los gerentes de control de calidad en algunos estados seleccionados aparecen enseguida.

Estado	Salario	Estado	Salario	Estado	Salario
Arizona	$35 000	Illinois	$58 000	Ohio	$50 000
California	49 100	Louisiana	60 000	Tennessee	60 000
Colorado	60 000	Maryland	60 000	Texas	71 400
Florida	60 000	Massachusetts	40 000	Virginia Oeste	60 000
Idaho	40 000	Nueva Jersey	65 000	Wyoming	55 000

Solución

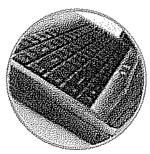
Un examen de los salarios revela que el salario anual de $60 000 se presenta con mayor frecuencia (seis veces) que otros salarios. Por tanto, la moda es $60 000.

En resumen, es posible determinar la moda para todos los niveles de datos, nominal, ordinal, de intervalo y de razón. La moda también tiene la ventaja de que no influyen en ella valores extremadamente grandes o pequeños.

Desventajas de la moda

No obstante, la moda tiene sus desventajas, por las cuales se le utiliza con menor frecuencia que a la media o a la mediana. En el caso de muchos conjuntos de datos no existe la moda, porque ningún valor se presenta más de una vez. Por ejemplo, no hay moda en el siguiente conjunto de datos de precios: \$19, \$21, \$23, \$20 y \$18. Sin embargo, como cada valor es diferente, podría argumentar que cada valor es la moda. Por lo contrario, en el caso de algunos conjuntos de datos hay más de una moda. Suponga que las edades de los miembros de un club de inversionistas son 22, 26, 27, 27, 31, 35 y 35. Ambas edades, 27 y 35 son modas. Así, este agrupamiento de edades se denomina *bimodal* (tiene dos modas). Alguien podría cuestionar la utilización de dos modas para representar la ubicación de este conjunto de datos de edades.

Autoevaluación 3.3



1. Una muestra de personas solteras en Towson, Texas, que reciben pagos por seguridad social reveló los siguientes subsidios mensuales: \$852, \$598, \$580, \$1 374, \$960, \$878 y \$1 130.
   a) ¿Cuál es la mediana del subsidio mensual?
   b) ¿Cuántas observaciones se encuentran debajo de la mediana? ¿Por encima de ella?
2. El número de interrupciones de trabajo en la industria automotriz en meses muestreados son de 6, 0, 10, 14, 8 y 0.
   a) ¿Cuál es la mediana en el número de interrupciones?
   b) ¿Cuántas observaciones se encuentran por debajo de la mediana? ¿Por encima de ella?
   c) ¿Cuál es el número modal de interrupciones de trabajo?

Ejercicios

17. ¿Qué informaría usted como valor modal para un conjunto de observaciones si hubiera un total de:
   a) 10 observaciones y no hubiera dos valores iguales?
   b) 6 observaciones, todas iguales?
   c) 6 observaciones con valores de 1, 2, 3, 4 y 4?

En los ejercicios 18 a 20, determine *a*) la media, *b*) la mediana y *c*) la moda.

18. Los siguientes son los números de cambios de aceite de los últimos 7 días en Jiffy Lube, que se ubica en la esquina de Elm Street y Pennsylvania Avenue.

41	15	39	54	31	15	33

19. El siguiente es el cambio porcentual en el ingreso neto de 2005 a 2006 en una muestra de 12 compañías de la construcción en Denver.

5	1	–10	–6	5	12	7	8	2	5	–1	11

20. Las siguientes son las edades de 10 personas en la sala de videojuegos del Southwyck Shopping Mall a las 10 de la mañana.

12	8	17	6	11	14	8	17	10	8

21. Abajo se enlistan diversos indicadores del crecimiento económico a largo plazo en Estados Unidos. Las proyecciones se extienden hasta el año 2008.

Indicador económico	Cambio porcentual	Indicador económico	Cambio porcentual
Inflación	4.5%	PNB real	2.9%
Exportaciones	4.7	Inversión (residencial)	3.6
Importaciones	2.3	Inversión (no residencial)	2.1
Ingreso real disponible	2.9	Productividad (total)	1.4
Consumo	2.7	Productividad (fabricación)	5.2

   a) ¿Cuál es la mediana del cambio porcentual?
   b) ¿Cuál es el cambio porcentual modal?

22. En la siguiente lista aparecen las ventas totales de automóviles (en millones de dólares) en Estados Unidos durante los pasados 14 años. En dicho periodo, ¿cuál fue la mediana en el número de automóviles vendidos? ¿Cuál es la moda?

9.0	8.5	8.0	9.1	10.3	11.0	11.5	10.3	10.5	9.8	9.3	8.2	8.2	8.5

23. La empresa de contabilidad de Rowatti y Koppel se especializa en la elaboración de declaraciones del impuesto sobre la renta de profesionales independientes, como médicos, dentistas, arquitectos y abogados. La firma emplea a 11 contadores que preparan declaraciones. El año pasado, el número de declaraciones elaboradas por cada contador fue la siguiente:

58	75	31	58	46	65	60	71	45	58	80

Determine la media, la mediana y la moda de los números de declaraciones elaboradas por cada contador. Si usted elaborara una, ¿qué medida de ubicación recomendaría que se presentara?

24. La demanda de videojuegos suministrados por Mid-Tech Video Games, Inc., se ha disparado en los últimos siete años. De ahí que el propietario requiera contratar técnicos que se mantengan a la par con la demanda. Mid-Tech proporciona a cada solicitante una prueba que el doctor McGraw, diseñador de la prueba, cree que se relaciona estrechamente con la habilidad para crear videojuegos. Para la población en general, la media de esta prueba es de 100. Enseguida aparecen los resultados de la prueba en el caso de lo aspirantes.

95	105	120	81	90	115	99	100	130	10

El presidente se encuentra interesado en las cualidades generales de los aspirantes al puesto basadas en la prueba. Calcule los resultados medio y mediano de los diez aspirantes. ¿Qué informaría usted al presidente? ¿Parece que los aspirantes son mejores que el resto de la población?

Solución con software

Con un paquete de software de estadística determine varias medidas de ubicación.

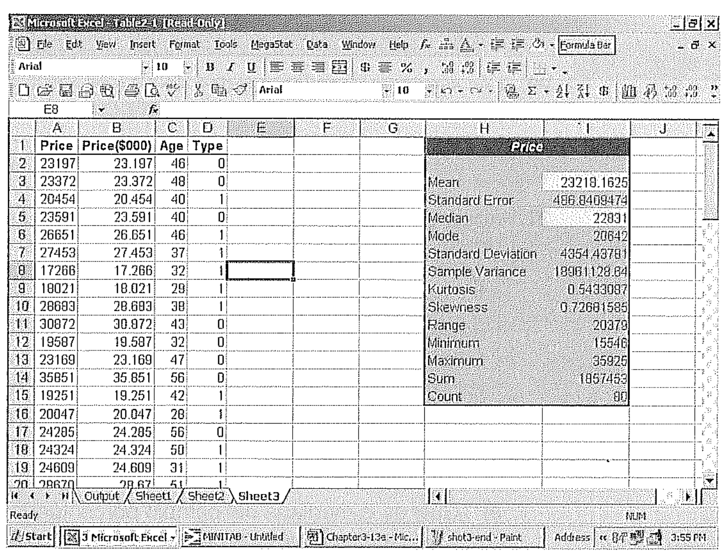
Ejemplo

La tabla 2.4 de la página 28 muestra los precios de 80 vehículos vendidos el mes pasado en Whitner Autoplex, en Raytown, Missouri. Determine los precios de venta medio y mediano.

Solución

Excel

Los precios de venta medio y mediano se presentan en el informe de la siguiente salida de Excel. (Recuerde que las instrucciones para crear la salida aparecen en la sección de **Comandos de software** localizada al final del capítulo.) En el estudio se incluyen 80 vehículos. Así que los cálculos con una calculadora resultarían tediosos y serían propensos a error.



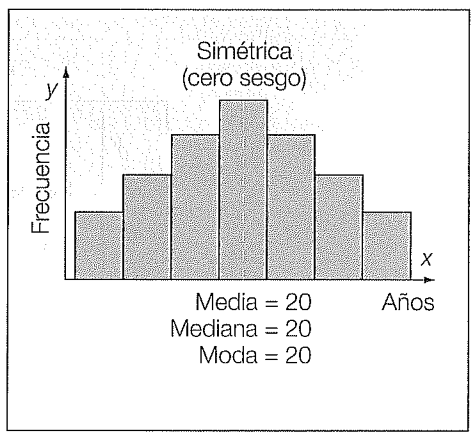
El precio promedio de ventas es de $23 218 y el mediano de $22 831. La diferencia entre estos dos valores es menor a $400. Así que cualquier valor es razonable. También es posible ver en la salida de Excel que se vendieron 80 vehículos, cuyo precio total es de $1 857 453. Más adelante se explicará el significado de error estándar, desviación estándar y otras medidas.

¿Qué podemos concluir? El precio de venta típico de un vehículo es de $23 000. La señora Ball de AutoUSA puede usar ese valor en la proyección de sus ingresos. Por ejemplo, si el representante puede incrementar el número de ventas en un mes, de 80 a 90, puede resultar un incremento en los ingresos de $230 000, encontrado por 10 × $23 000.

Posiciones relativas de la media, la mediana y la moda

En una distribución en forma de campana la media, la mediana y la moda son iguales

Observe el histograma de la figura 3.2. Se trata de una distribución simétrica que también tiene forma de campana. Esta distribución *posee la misma forma a cualquier lado del centro.* Si el polígono estuviera doblado a la mitad, las dos mitades serían idénticas. En cualquier distribución simétrica la moda, la mediana y la media siempre son iguales. Son equivalentes a 20 años en la gráfica 3.2. Hay distribuciones simétricas que no tienen forma de campana.



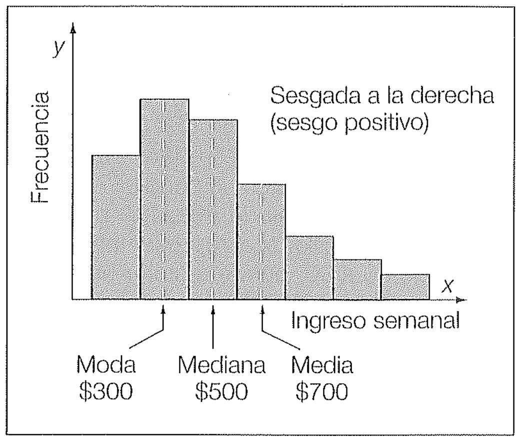
GRÁFICA 3.2 Distribución simétrica

El número de años correspondiente al punto más alto de la curva es la *moda* (20 años). Como la distribución es simétrica, la *mediana* corresponde al punto en el que la distribución se divide a la mitad (20 años). El número total de frecuencias que representan muchos años se encuentra compensado por el número total que representa pocos años, lo cual da como resultado una *media aritmética* de 20 años. Cualquiera de estas tres medidas sería adecuada para representar el centro de la distribución.

Una distribución sesgada no es simétrica

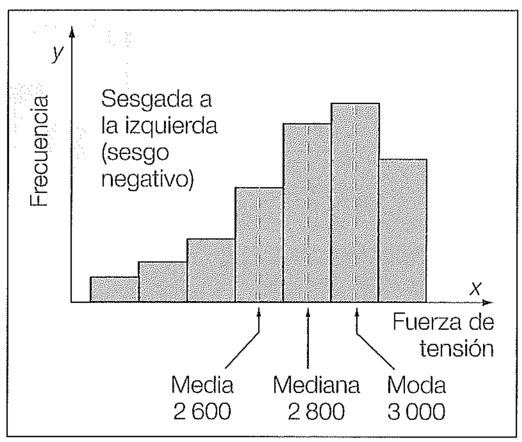
Si una distribución no es simétrica, o **sesgada**, la relación entre las tres medidas cambia. En una **distribución con sesgo positivo** la media aritmética es la mayor de las tres medidas. ¿Por qué? En ella influyen más que sobre la mediana o la moda unos cuantos valores extremadamente altos. La mediana es, por lo general, la siguiente medida más grande en una distribución de frecuencias con sesgo positivo. La moda es la menor de las tres medidas.

Si la distribución tiene un sesgo muy pronunciado, como en el caso de los ingresos semanales de la gráfica 3.3, la media no sería una medida adecuada. La mediana y la moda serían más representativas.



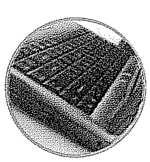
GRÁFICA 3.3 Distribución con sesgo positivo

Por lo contrario, si una distribución tiene un **sesgo negativo**, la media es la menor medida de las tres. Por supuesto, la media es sensible a la influencia de una cantidad extremadamente pequeña de observaciones. La mediana es mayor que la media aritmética y la moda es la más grande de las tres medidas. De nuevo, si la distribución tiene un sesgo muy pronunciado, como la distribución de fuerzas de tensión que se muestran en la gráfica 3.4, la media no se utilizaría para representar a los datos.



GRÁFICA 3.4 Distribución con sesgo negativo

Autoevaluación 3.4



Las ventas semanales de una muestra de tiendas de suministros electrónicos de alta tecnología se organizaron en una distribución de frecuencias. La media de las ventas semanales que se calculó fue de $105 900, la mediana de $105 000 y la moda de $104 500.

a) Trace una gráfica de las ventas con la forma de un polígono de frecuencias suavizado. Observe la ubicación de la media, la mediana y la moda sobre el eje X.

b) ¿La distribución es simétrica, tiene un sesgo positivo o un sesgo negativo? Explique su respuesta.

Ejercicios

25. La tasa de desempleo en el estado de Alaska durante los 12 meses de 2004 aparece en la siguiente tabla:

Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dic
8.7	8.8	8.7	7.8	7.3	7.8	6.6	6.5	6.5	6.8	7.3	7.6

a) ¿Cuál es la media aritmética para la tasa de desempleo en Alaska?
b) Encuentre la media y la moda para la tasa de desempleo.
c) Calcule la media aritmética y la mediana sólo para los meses de invierno (de diciembre a marzo). ¿Es muy diferente?

26. Big Orange Trucking diseña un sistema de información que se utiliza para comunicaciones *en cabina*. Debe resumir datos de ocho sitios de cierta zona para describir condiciones típicas. Calcule una medida adecuada de ubicación central para cada una de las tres variables que aparecen en la siguiente tabla:

Ciudad	Dirección del viento	Temperatura	Pavimento
Anniston, AL	Oeste	89	Seco
Atlanta, GA	Noroeste	86	Mojado
Augusta, GA	Suroeste	92	Mojado
Birmingham, AL	Sur	91	Seco
Jackson, MS	Suroeste	92	Seco
Meridian, MS	Sur	92	Sendero
Monroe, LA	Suroeste	93	Mojado
Tuscaloosa, AL	Suroeste	93	Sendero

Media geométrica

La media geométrica nunca es mayor que la media aritmética

La media geométrica resulta útil para determinar el cambio promedio de porcentajes, razones, índices o tasas de crecimiento. Posee amplias aplicaciones en la administración y la economía, ya que con frecuencia hay interés en determinar los cambios porcentuales de ventas, salarios o cifras económicas, como el producto interno bruto, los cuales se combinan o se basan unos en otros. La media geométrica de un conjunto de n números positivos se define como la raíz enésima de un producto de n variables. La fórmula de la media geométrica se escribe de la siguiente manera:

MEDIA GEOMÉTRICA $$GM = \sqrt[n]{(X_1)(X_2)\cdots(X_n)} \qquad \textbf{[3.4]}$$

La media geométrica siempre es menor o igual (nunca mayor que) que la media aritmética. Todos los datos deben ser positivos.

Como ejemplo de media geométrica, asuma que usted recibe 5% de incremento en el salario este año y 15% de incremento el siguiente. El incremento porcentual anual promedio es de 9.886, no de 10. ¿Por qué razón? Comience calculando la media geométrica. Recuerde, por ejemplo, que 5% de incremento salarial equivale a 105%. Lo que expresa como 1.05.

$$GM = \sqrt{(1.05)(1.15)} = 1.09886$$

Este resultado puede verificarse suponiendo que su ingreso mensual fue de \$3 000 para comenzar y que recibió dos incrementos de 5% y 15%.

$$\begin{aligned} \text{Incremento } 1 &= \$3\,000(.05) = \$150.00 \\ \text{Incremento } 2 &= \$3\,150(.15) = \underline{\ \ 472.50} \\ \text{Total} &\qquad\qquad\quad \$622.50 \end{aligned}$$

El incremento total a su salario es de \$622.50. Esto equivale a:

$$\$3\,000(.09886) = \$296.58$$
$$\$3\,150(.09886) = 325.90$$
$$\$622.48 \text{ es de alrededor de } \$622.50$$

El siguiente ejemplo muestra la media geométrica de diversos porcentajes.

Ejemplo

La recuperación de una inversión realizada por Atkins Construction Company durante cuatro años consecutivos fue de 30%, 20%, –40% y 200%. ¿Cuál es la media geométrica de la recuperación de la inversión?

Solución

El número 1.3 representa 30% de la recuperación de la inversión, que es la inversión *original* de 1.0 más la *recuperación* de 0.3. El número 0.6 representa la pérdida de 40%, que es la inversión original de 1.0 menos la pérdida de 0.4. Este cálculo supone que el total de la inversión de cada periodo se reinvierte o se convierte en la base de la siguiente. En otras palabras, la base para el segundo periodo es 1.3 y la base para el tercer periodo es (1.3)(1.2) y así sucesivamente.

Entonces la media geométrica de la tasa de recuperación es de 29.4%, que se determina por medio del siguiente cálculo:

$$GM = \sqrt[n]{(X_1)(X_2)\cdots(X_n)} = \sqrt[4]{(1.3)(1.2)(0.6)(3.0)} = \sqrt[4]{2.808} = 1.294$$

De esta manera, la media geométrica es la raíz cuarta de 2.808. Así, la tasa promedio de recuperación (tasa de crecimiento anual compuesta) es de 29.4%.

Observe, asimismo, que si calcula la media aritmética [(30 + 20 – 40 + 200)/4 = 52.5], obtendrá un número mucho más grande, lo que dispararía la tasa de recuperación real.

Otro modelo de aplicación de la media geométrica tiene que ver con determinar un cambio porcentual promedio durante cierto periodo. Por ejemplo, si usted ganó \$30 000 en 1997 y \$50 000, en 2007, ¿cuál es la tasa anual de incremento durante el periodo? Ésta es de 5.24%. La tasa de incremento se determina a partir de la siguiente fórmula.

PORCENTAJE PROMEDIO QUE SE INCREMENTA CON EL TIEMPO

$$GM = \sqrt[n]{\frac{\text{Valor al final del periodo}}{\text{Valor al inicio del periodo}}} - 1 \qquad \textbf{[3.5]}$$

En el recuadro anterior n es el número de periodos. Un ejemplo mostrará los detalles para determinar el incremento porcentual anual.

Ejemplo

Durante la década de los noventa y hasta los primeros años del 2000, Las Vegas, Nevada, fue la ciudad de mayor crecimiento en Estados Unidos. La población se incrementó de 258 295 en 1990 a 534 847 en 2005. Es un incremento de 276 552 personas o 107% de incremento durante el periodo de 15 años. ¿Cuál es el incremento *anual* promedio?

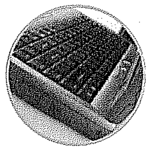
Solución

Hay 15 años entre 1990 y 2005, así que n = 15. De esta manera, la fórmula 3.5 de la media geométrica, aplicada a este problema, se transforma en:

$$GM = \sqrt[n]{\frac{\text{Valor al final de periodo}}{\text{Valor al inicio del periodo}}} - 1.0 = \sqrt[15]{\frac{534\,847}{258\,295}} - 1.0 = 1.0497 - 1.0 = .0497$$

El valor de 0.0497 indica que el crecimiento anual promedio durante el periodo de 15 años fue de 4.97%. Expresado en otros términos, la población de Las Vegas creció a una tasa de 4.97% por año de 1990 a 2005.

Autoevaluación 3.5



1. El incremento porcentual en ventas de los pasados 4 años en Combs Cosmetics fue de 4.91, 5.75, 8.12 y 21.60.
   a) Determine la media geométrica del incremento porcentual.
   b) Determine la media aritmética del incremento porcentual.
   c) ¿Es igual la media aritmética a la media geométrica o mayor?
2. La producción de camiones Cablos se elevó de 23 000 unidades en 1996 a 120 520 unidades en 2006. Calcule la media geométrica del incremento porcentual anual.

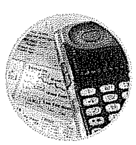
Ejercicios

27. Calcule la media geométrica de los siguientes incrementos porcentuales: 8, 12, 14, 26 y 5.
28. Estime la media geométrica de los siguientes incrementos porcentuales: 2, 8, 6, 4 10, 6, 8 y 4.
29. A continuación se enlista el incremento porcentual en ventas de MG Corporation para los pasados 5 años. Determine la media geométrica del incremento porcentual en ventas durante el periodo.

9.4	13.8	11.7	11.9	14.7

30. En 1996 un total de 14 968 000 contribuyentes en Estados Unidos presentaron en forma electrónica sus declaraciones de impuestos. Para el año 2004 el número se había incrementado a 66 290 000. ¿Cuál es la media geométrica del incremento anual para el periodo?
31. El U.S. Bureau of Labor Statistics publica mensualmente el índice de precios al consumidor. Informa el cambio de precios en una canasta de artículos en el mercado de un periodo a otro. El índice para 1994 fue de 148.2, para 2004 se incrementó a 188.9 ¿Cuál es la media geométrica del incremento anual de dicho periodo?
32. En 1976 el precio promedio en Estados Unidos de un galón de gasolina sin plomo en una estación de autoservicio era de $0.605. Para el año 2005, el precio promedio se había incrementado a $2.57. ¿Cuál es la media geométrica del incremento anual en dicho periodo?
33. En 2001 había 42 millones de suscriptores al servicio de buscapersonas. Para el año 2006 el número de suscriptores aumentó a 70 millones. ¿Cuál es la media geométrica del incremento anual de dicho periodo?
34. La información que sigue muestra el costo de un año de estudios en universidades públicas y privadas en 1992 y 2004. ¿Cuál es la media geométrica del incremento anual en dicho periodo en el caso de las dos clases de escuelas? Compare las tasas de incremento.

Tipo de universidad	1992	2004
Pública	$ 4 975	$ 11 354
Privada	12 284	27 516



Estadística en acción

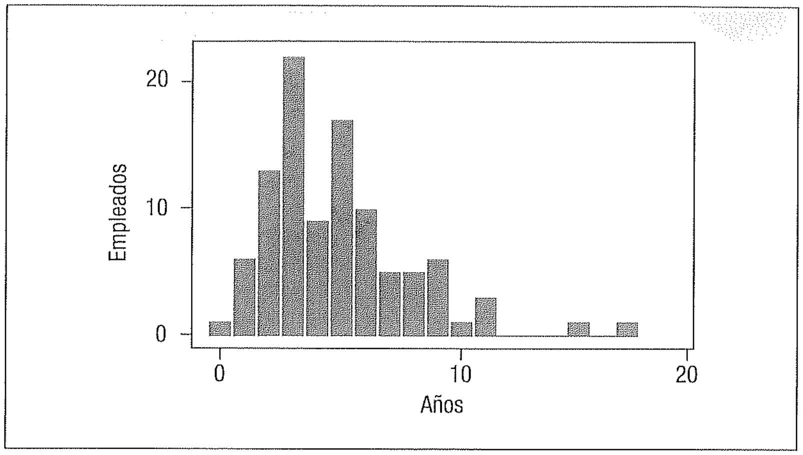
El servicio postal de Estados Unidos ha intentado comportarse de forma más *amigable* con el usuario en los últimos siete años. Una encuesta reciente mostró que los consumidores estaban interesados en que hubiera más regularidad en los tiempos de entrega. Antes una carta local podría tardar en llegar un día o varios. "Sólo díganme con cuántos días de anticipación tengo que enviar una tarjeta de felicitación a mamá para que llegue el día de su cumpleaños, ni antes ni después", era una queja común. El nivel de regularidad se mide a partir de la desviación estándar de los tiempos de entrega.

¿Por qué estudiar la dispersión?

Una medida de ubicación, como la media o la mediana, solamente describe el centro de los datos. Desde este punto de vista resulta valiosa, pero no dice nada sobre la dispersión de los datos. Por ejemplo, si la guía de turismo ecológico dice que el río que se encuentra adelante tiene en promedio 3 pies de profundidad, ¿querría usted cruzarlo a pie sin más información? Quizá no. Usted desearía saber algo sobre la variación de la profundidad. ¿Mide 3.25 pies la máxima profundidad y 2.75 pies la mínima? En dicho caso, usted estaría de acuerdo en cruzar. ¿Qué hay si usted se enteró de que la profundidad del río variaba de 0.50 pies a 5.5 pies? Su decisión probablemente sería no cruzar. Antes de tomar una decisión sobre cruzar el río, usted desea información tanto de la profundidad típica como de la dispersión de la profundidad del río.

Un valor pequeño en una medida de dispersión indica que los datos se acumulan con proximidad alrededor de la media aritmética. Por consiguiente, la media se considera representativa de los datos. Por lo contrario, una medida grande de dispersión indica que la media no es confiable (vea la gráfica 3.5). Los 100 empleados de Hammond Iron

Works, Inc., una compañía que fabrica acero, se organizan en un histograma basado en el número de años que los empleados han laborado en la compañía. La media de 4.9 años no es muy representativa de los empleados.

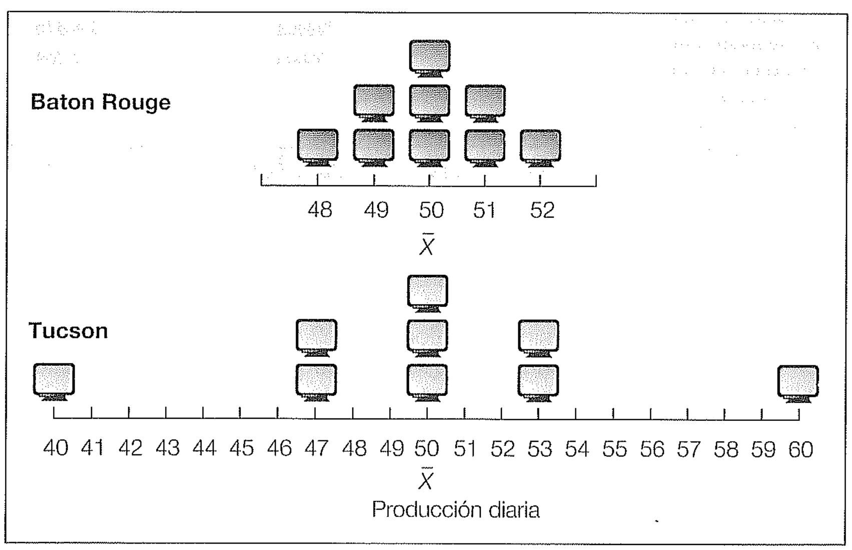


GRÁFICA 3.5 Histograma de los años laborados para Hammond Iron Works, Inc.

El promedio no es representativo como consecuencia de que la dispersión es grande

Una segunda razón para estudiar la dispersión en un conjunto de datos consiste en comparar la propagación en dos o más distribuciones. Por ejemplo, asuma que el nuevo monitor de computadora Vision Quest LCD se arma en Baton Rouge y también en Tucson. La producción media aritmética por hora tanto en la planta de Baton Rouge como en la de Tucson es de 50. Sobre la base de las dos medias, podría concluir que las distribuciones de las producciones por hora son idénticas. Sin embargo, los registros de producción de 9 horas en las dos plantas revelan que esta conclusión no es correcta (vea la gráfica 3.6). La producción de Baton Rouge varía de 48 a 52 montajes por hora. La producción en la planta de Tucson es más errática, ya que varía de 40 a 60 la hora. Por tanto, la producción por hora en Baton Rouge se acumula cerca de la media de 50; la producción por hora de Tucson es más dispersa.

Una medida de dispersión sirve para evaluar la confiabilidad de dos o más medidas de ubicación



GRÁFICA 3.6 Producción por hora de monitores de computadora en las plantas de Baton Rouge y Tucson

Medidas de dispersión

Consideraremos diversas medidas de dispersión. El rango se sustenta en los valores máximo y mínimo del conjunto de datos. La desviación media, la varianza y la desviación estándar se basan en desviaciones de la media aritmética.

Rango

La medida más simple de dispersión es el **rango**. Representa la diferencia entre los valores máximo y mínimo de un conjunto de datos. En forma de ecuación:

RANGO Rango = Valor máximo – valor mínimo **[3.6]**

El rango se emplea mucho en aplicaciones de control de procesos estadísticos (CPE) como consecuencia de que resulta fácil de calcular y entender.

Ejemplo

Consulte la gráfica 3.6. Determine el rango del número de monitores de computadora producidos por hora en las plantas de Baton Rouge y Tucson. Interprete los dos rangos.

Solución

El rango de la producción por hora de monitores de computadora en la planta de Baton Rouge es de 4, el cual se determina por la diferencia entre la producción máxima por hora de 52 y la mínima de 48. El rango de la producción por hora en la planta de Tucson es de 20 monitores de computadora, obtenido con el cálculo 60 – 40. Por tanto: 1. Existe menos dispersión en la producción por hora en la planta de Baton Rouge que en la planta de Tucson, porque el rango de 4 monitores de computadora es menor que el rango de 20 monitores; 2. La producción se acumula más alrededor de la media de 50 en la planta de Baton Rouge que en la planta de Tucson (ya que un rango de 4 es menor que un rango de 20). Así, la producción media en la planta de Baton Rouge (50 monitores de computadora) resulta una medida de ubicación más representativa que la media de 50 monitores de computadora en la planta de Tucson.

Desviación media

Un problema que presenta el rango estriba en que parte de dos valores, el más alto y el más bajo; no toma en cuenta todos los valores. La **desviación** media sí lo hace; mide la cantidad media respecto de la cual los valores de una población o muestra varían. Expresado esto en forma de definición:

DESVIACIÓN MEDIA Media aritmética de los valores absolutos de las desviaciones con respecto a la media aritmética.

En el caso de una muestra, la desviación media, designada *DM,* se calcula mediante la fórmula:

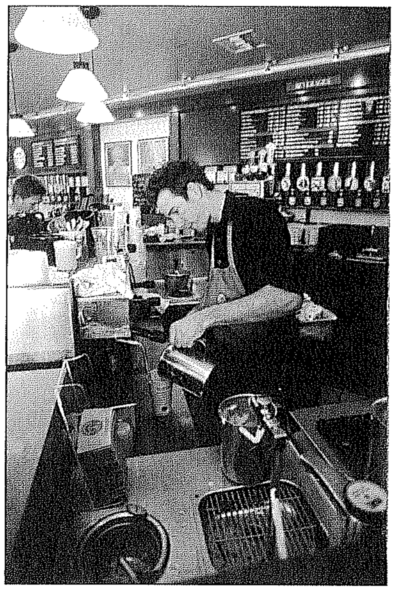
DESVIACIÓN MEDIA
$$MD = \frac{\Sigma|X - \bar{X}|}{n} \qquad \textbf{[3.7]}$$

en la cual:

X es el valor de cada observación;
$\bar{X}$ es la media aritmética de los valores;
n es el número de observaciones en la muestra;
$| \, |$ indica el valor absoluto.

¿Por qué ignorar los signos de las desviaciones de la media? De no hacerlo las desviaciones positivas y negativas de la media se compensarían con exactitud unas a otras y la desviación media siempre sería cero. Dicha medida (cero) resultaría un estadístico sin utilidad.

Ejemplo



El número de capuchinos vendidos en el local de Starbucks de Orange County Airport entre las cuatro y las siete de la tarde de una muestra de 5 días el año pasado fue de 20, 40, 50, 60 y 80. En el aeropuerto de LAX en Los Ángeles, el número de capuchinos vendidos en el local de Starbucks entre las cuatro y la siete de la tarde de una muestra de 5 días el año pasado fue de 20, 49, 50, 51 y 80. Determine la media, la mediana, el rango y la desviación media de cada local. Compare las diferencias.

Solución

En el caso del local de Orange County, la media, la mediana y el rango son:

Media	50 capuchinos por día
Mediana	50 capuchinos por día
Rango	60 capuchinos por día

La desviación media es la media de las diferencias entre las observaciones individuales y la media aritmética. En el caso de Orange County, la cantidad media de capuchinos vendida es de 50, el cálculo es (20 + 40 + 50 + 80)/5. Enseguida determine las diferencias entre cada observación y la media. Enseguida sume estas diferencias, haciendo caso omiso de los signos, y divida la suma entre el número de observaciones. El resultado es la diferencia media entre las observaciones y la media.

Número de observaciones	$(X - \bar{X})$	Desviación absoluta
20	(20 − 50) = −30	30
40	(40 − 50) = −10	10
50	(50 − 50) = 0	0
60	(60 − 50) = 10	10
80	(80 − 50) = 30	30
	Total	80

$$MD = \frac{\Sigma|X - \bar{X}|}{n} = \frac{80}{5} = 16$$

La desviación media es de 16 capuchinos al día: el número de capuchinos vendidos se desvía, en promedio, 16 unidades de la media de 50 capuchinos al día.

En seguida aparece el resumen de la media, la mediana, el rango y la desviación media en el caso de LAX. Realice los cálculos para verificar los resultados.

Media	50 capuchinos por día
Mediana	50 capuchinos por día
Rango	60 capuchinos por día
Desviación media	12.4 capuchinos por día

Recuerde que en el capítulo anterior se le describieron datos mediante métodos gráficos. En este capítulo se emplearán medidas numéricas para describirlos. Cuando emplee medidas numéricas, es muy importante informar siempre las medidas de ubicación y de dispersión.

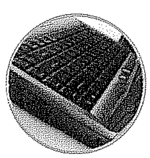
Interprete y compare los resultados de las medidas en el caso de las tiendas de Starbucks. La media y la mediana de las dos tiendas son exactamente las mismas, 50 capuchinos al día. Por consiguiente, la ubicación de ambas distribuciones es la misma. El rango en ambas tiendas también es el mismo, 60. Sin embargo, recuerde que el rango proporciona información limitada sobre la dispersión de la distribución.

Observe que las desviaciones medias no son las mismas porque se basan en las diferencias entre todas las observaciones y la media aritmética, que muestra la relativa proximidad o acumulación de los datos concerniente a la media o centro de la distribución. Compare la desviación media de Orange County de 16 con la desviación de LAX de 12.4. Sobre la base de la desviación media, es posible decir que la dispersión de la distribución de ventas de LAX Starbucks se encuentra más concentrada cerca de la media de 50 que en la tienda de Orange County.

Ventajas de la desviación media

La desviación media posee dos ventajas. Primero, incluye todos los valores de los cálculos. Recuerde que el rango sólo incluye los valores máximo y mínimo. Segundo, es fácil de definir: es la cantidad promedio que los valores se desvían de la media. Sin embargo, su inconveniente es el empleo de valores absolutos. Por lo general, es difícil trabajar con valores absolutos, así que la desviación media no se emplea con tanta frecuencia como otras medidas de dispersión, como la desviación estándar.

Autoevaluación 3.6



Los pesos de los contenedores enviados a Irlanda son (en miles de libras):

95	103	105	110	104	105	112	90

a) ¿Cuál es el rango de los pesos?
b) Calcule el peso medio aritmético.
c) Estime la desviación media de los pesos.

Ejercicios

En los ejercicios 35-38, calcule: *a*) el rango; *b*) la media aritmética; *c*) la desviación media; *d*) el rango. Interprete los valores que obtenga.

35. Hubo cinco representantes de servicio al cliente trabajando en Electronic Super Store durante la pasada venta de fin de semana. Las cantidades de HDTV que vendieron estos representantes son: 5, 8, 4, 10 y 3.

36. El Departamento de Estadística de la Western State University ofrece ocho secciones de estadística básica. En seguida aparecen los números de estudiantes matriculados en estas secciones: 34, 46, 52, 29, 41, 38, 36 y 28.

37. Dave´s Automatic Door instala puertas automáticas para cocheras. La siguiente lista indica el número de minutos que se requieren para instalar una muestra de 10 puertas automáticas: 28, 32, 24, 46, 44, 40, 54, 38, 32 y 42.

38. Una muestra de ocho compañías de la industria aeronáutica participaron en una encuesta sobre la recuperación de la inversión que tuvieron el año pasado. Los resultados (en porcentaje) son los siguientes: 10.6, 12.6, 14.8, 18.2, 12.0, 14.8, 12.2 y 15.6.

39. Diez adultos jóvenes que viven en California, elegidos al azar, calificaron el sabor de una nueva pizza de sushi con atún, arroz y kelp en una escala de 1 a 50, en la que el 1 indica que no les gusta el sabor y 50 que sí les gusta. Las calificaciones fueron las siguientes:

34	39	40	46	33	31	34	14	15	45

En un estudio paralelo 10 adultos jóvenes, elegidos al azar, en Iowa calificaron el sabor de la misma pizza. Las calificaciones fueron las siguientes:

28	25	35	16	25	29	24	26	17	20

Como investigador de mercado, compare los mercados potenciales para la pizza de sushi.

40. Una muestra de archivos de personal de ocho empleados en las instalaciones de Pawnee de Acme Carpet Cleaners, Inc., reveló que durante el último semestre éstos perdieron la siguiente cantidad de días por enfermedad:

2	0	6	3	10	4	1	2

Durante el mismo periodo, una muestra de ocho empleados en las instalaciones de Chickpee de Acme Carpets reveló que ellos perdieron las siguientes cantidades de días por enfermedad:

2	0	1	0	5	0	1	0

Como director de relaciones humanas, compara las dos instalaciones. ¿Qué recomendaría?

Varianza y desviación estándar

La varianza y la desviación estándar se basan en las desviaciones de la media elevadas al cuadrado

La **varianza** y la **desviación estándar** también se fundamentan en las desviaciones de la media. Sin embargo, en lugar de trabajar con el valor absoluto de las desviaciones, la varianza y la desviación estándar lo hacen con el cuadrado de las desviaciones.

VARIANZA Media aritmética de las desviaciones de la media elevadas al cuadrado.

La varianza es no negativa y es cero sólo si todas las observaciones son las mismas.

DESVIACIÓN ESTÁNDAR Raíz cuadrada de la varianza.

Varianza de la población Las fórmulas de la varianza poblacional y la varianza de la muestra son ligeramente diferentes. La varianza de la población se estudia primero. (Recuerde que una población es la totalidad de las observaciones estudiadas.) La **varianza de la población** se determina de la siguiente manera:

VARIANZA DE LA POBLACIÓN $$\sigma^2 = \frac{\Sigma(X-\mu)^2}{N}$$ **[3.8]**

En esta fórmula:

σ^2 es la varianza de la población (σ es la letra minúscula griega sigma); se lee *sigma al cuadrado;*

X es el valor de una observación de la población;

μ es la media aritmética de la población;

N es el número de observaciones de la población.

Observe el proceso de cálculo de la varianza:

- Comience determinando la media;
- En seguida calcule la diferencia entre cada observación y la media, y eleve al cuadrado dicha diferencia;
- Entonces sume todas las diferencias elevadas al cuadrado;
- Por último divida la suma de las diferencias elevadas al cuadrado entre el número de elementos de la población.

Así, usted podría pensar que la varianza de la población es la media de las diferencias elevadas al cuadrado entre cada valor y la media. En las poblaciones cuyos valores cercanos a la media, la varianza de la población puede ser pequeña.En las poblaciones cuyos valores se apartan de la media, la varianza de la población puede ser grande.

La varianza compensa el inconveniente que presenta el rango gracias a los valores absolutos de la población, mientras que el rango incluye sólo los valores máximo y mínimo. El problema de que $\Sigma(X - \mu) = 0$, se corrige elevando al cuadrado las diferencias, en lugar de emplear valores absolutos. Elevar al cuadrado las diferencias siempre dará como resultado valores no negativos.

Ejemplo

El número de multas de tránsito levantadas durante los pasados cinco meses en Beaufort County, Carolina del Sur, es de 38, 26, 13, 41 y 22. ¿Cuál es la varianza de la población?

Solución

Número (X)	$X - \mu$	$(X - \mu)^2$
38	+10	100
26	−2	4
13	−15	225
41	+13	169
22	−6	36
140	0*	534

$$\mu = \frac{\Sigma X}{N} = \frac{140}{5} = 28$$

$$\sigma^2 = \frac{\Sigma(X-\mu)^2}{N} = \frac{534}{5} = 106.8$$

*La suma de las desviaciones de la media debe ser igual a cero.

Como en el caso del rango y la desviación media, la varianza se emplea para comparar la dispersión en dos o más conjuntos de observaciones. Por ejemplo, se calculó que la varianza del número de multas levantadas en Beaufort County fue de 106.8. Si la varianza del número de multas levantadas en Marlboro County, Carolina del Sur, es de 342.9, se concluye que: 1. Hay menos dispersión en la distribución del número de multas levantadas en Beaufort (ya que 106.8 es menor que 342.9); 2. El número de multas levantadas en Beaufort County se encuentran más apiñadas en torno a la media de 28 que el número de multas levantadas en Marlboro County. Por consiguiente, la media de multas levantadas en Beaufort County constituye una medida de ubicación más representativa que la media de multas en Marlboro County.

La varianza resulta difícil de interpretar porque las unidades se elevan al cuadrado

Desviación estándar de la población Tanto el rango como la desviación media resultan fáciles de interpretar. El rango es la diferencia entre los valores alto y bajo de un conjunto de datos, y la desviación media es la media de las desviaciones de la media.

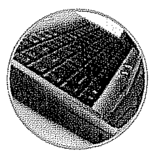
Sin embargo, la varianza resulta difícil de interpretar en el caso de un solo conjunto de observaciones. La varianza de 106.8 del número de multas levantadas no se expresa en términos de multas, sino de multas elevadas al cuadrado.

La desviación estándar se expresa en las mismas unidades de los datos

Existe una forma de salir del problema. Si extrae la raíz cuadrada de la varianza de la población, puede convertirla a las mismas unidades de medición empleadas en los datos originales. La raíz cuadrada de 106.8 multas elevadas al cuadrado es de 10.3 multas. Las unidades ahora son sencillamente multas. La raíz cuadrada de la varianza de la población es la **desviación estándar de la población**.

DESVIACIÓN ESTÁNDAR DE LA POBLACIÓN $$\sigma = \sqrt{\frac{\Sigma(X-\mu)^2}{N}} \qquad [3.9]$$

Autoevaluación 3.7



Este año la oficina en Filadelfia de Price Waterhouse Coopers LLP contrató a cinco contadores que están haciendo prácticas. Los salarios mensuales iniciales de éstos fueron de \$3 536, \$3 173, \$3 448, \$3 121 y \$3 622.

a) Calcule la media de la población.
b) Estime la varianza de la población.
c) Aproxime la desviación estándar de la población.
d) La oficina de Pittsburgh contrató a cinco empleados que están haciendo prácticas. El salario mensual promedio fue de \$3 550 y la desviación estándar de \$250. Compare los dos grupos.

Ejercicios

41. Considere en una población los siguientes cinco valores: 8, 3, 7, 3 y 4.
a) Determine la media de la población.
b) Determine la varianza.

42. Considere a los siguientes seis valores como una población: 13, 3, 8, 10, 8 y 6.
a) Determine la media de la población.
b) Determine la varianza.

43. El informe anual de Dennis Industries incluyó las siguientes ganancias primarias por acción común durante los pasados 5 años: \$2.68, \$1.03, \$2.26, \$4.30 y \$3.58. Si supone que éstos son los valores poblacionales,
a) ¿Cuáles son las medias aritméticas de las ganancias primarias por acción común?
b) ¿Cuál es la varianza?

44. Con respecto al ejercicio 43, el informe anual de Dennis Industries también arrojó estos rendimientos sobre valores de renta variable para el mismo periodo de cinco años (en porcentaje): 13.2, 5.0, 10.2, 17.5 y 12.9.
a) ¿Cuál es la media aritmética del rendimiento?
b) ¿Cuál es la varianza?

45. Plywood, Inc., informó las siguientes utilidades sobre valores de renta variable durante los pasados 5 años: 4.3, 4.9, 7.2, 6.7 y 11.6. Considere estos valores como poblacionales.
a) Calcule el rango, la media aritmética, la varianza y la desviación estándar.
b) Compare las utilidades sobre valores de renta variable de Playwood, Inc., con las de Dennis Industries citadas en el ejercicio 44.

46. Los ingresos anuales de cinco vicepresidentes de TMV Industries son: \$125 000, \$128 000, \$122 000, \$133 000 y \$140 000. Considere estos valores como una población.
a) ¿Cuál es el rango?
b) ¿Cuál es el ingreso medio aritmético?
c) ¿Cuál es la varianza poblacional? ¿La desviación estándar?
d) También se estudiaron los ingresos anuales de personal de otra empresa similar a TMV. La media fue de \$129 000 y la desviación estándar de \$8 612. Compare las medias y dispersiones de las dos firmas.

Varianza muestral La fórmula para la media poblacional es $\mu = \Sigma X/N$. Sencillamente cambie los símbolos para la media de la muestra; es decir, $\bar{X} = \Sigma X/n$. Por desgracia, la conversión de una varianza poblacional en una varianza muestral no es tan directa.

Requiere un cambio en el denominador. En lugar de sustituir n (el número en la muestra) por N (el número en la población), el denominador es $n - 1$. Así, la fórmula de la **varianza muestral** es:

VARIANZA MUESTRAL $$s^2 = \frac{\Sigma(X - \bar{X})^2}{n-1} \quad [3.10]$$

en la cual:

- s^2 es la varianza muestral;
- X es el valor de cada observación de la muestra;
- $\bar{X}$ es la media de la muestra;
- n es el número de observaciones en la muestra.

¿Por qué se hizo este cambio en el denominador? Aunque el empleo de n se entiende en virtud de que se utiliza $\bar{X}$ para calcular μ, esto tiende a subestimar la varianza poblacional, σ^2. La inclusión de $(n - 1)$ en el denominador proporciona la corrección adecuada para esta tendencia. Como la aplicación fundamental de estadísticos muestrales como s^2 es calcular parámetros de población como σ^2, se prefiere $(n - 1)$ en lugar de n para definir la varianza muestral. También se emplea esta convención al calcular la desviación estándar de una muestra.

Ejemplo

Los salarios por hora de una muestra de empleados de medio tiempo de Home Depot son: \$12, \$20, \$16, \$18 y \$19. ¿Cuál es la varianza de la muestra?

Solución

La varianza muestral se calcula con la fórmula 3.10.

$$\bar{X} = \frac{\Sigma X}{n} = \frac{\$85}{5} = \$17$$

Salario por hora (X)	$X - \bar{X}$	$(X - \bar{X})^2$
\$12	−\$5	25
20	3	9
16	−1	1
18	1	1
19	2	4
\$85	0	40

$$s^2 = \frac{\Sigma(X - \bar{X})^2}{n-1} = \frac{40}{5-1}$$

$$= 10 \text{ en dólares al cuadrado}$$

Desviación estándar de la muestra La desviación estándar de la muestra se utiliza como estimador de la desviación estándar de la población. Como se hizo notar, la desviación estándar de la población es la raíz cuadrada de la varianza de la población. Asimismo, la *desviación estándar de la muestra es la raíz cuadrada de la varianza de la muestra.* La desviación estándar de la muestra se calcula con mayor facilidad de la siguiente manera:

DESVIACIÓN ESTÁNDAR DE LA MUESTRA $$s = \sqrt{\frac{\Sigma(X - \bar{X})^2}{n-1}} \quad [3.11]$$

Ejemplo

La varianza de la muestra en el ejemplo anterior, que incluye salarios por hora, se calculó en 10. ¿Cuál es la desviación estándar?

Solución

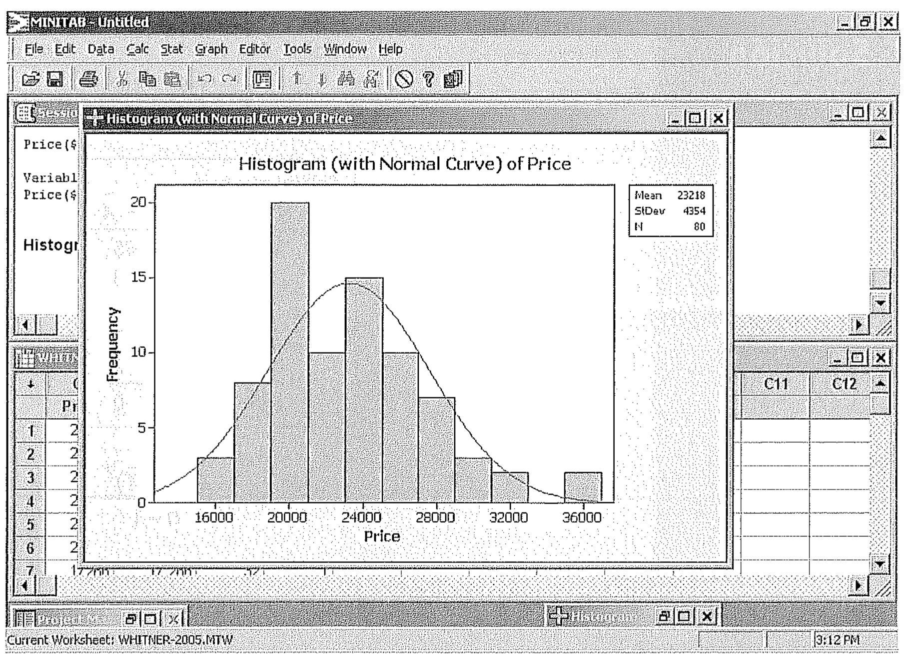
La desviación estándar de la muestra es \$3.16, que se determina con $\sqrt{10}$. Observe nuevamente que la varianza de la muestra se expresa en términos de dólares al cuadrado, pero al extraer la raíz cuadrada a 10 se obtiene \$3.16, que se encuentra en las mismas unidades (dólares) que los datos originales.

Solución con software

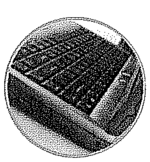
En la página 66 utilizamos Excel para determinar la media y la mediana de los datos de ventas de Whitner Autoplex. También notará que Excel presenta la desviación estándar de la muestra. Como la mayoría de los paquetes de software de estadística, Excel supone que los datos corresponden a una muestra.

Otro paquete de software que utilizará en el libro es MINITAB. El paquete utiliza un formato de hoja de cálculo, muy parecido a Excel, aunque genera una variedad más amplia de datos de estadística. Enseguida aparece la información de los precios de venta de Whitner Autoplex. Observe que se incluye un histograma (aunque la acción predeterminada consiste en utilizar un intervalo de clase de \$2 000 con 11 clases), así como la media, la desviación estándar de la muestra y el número de observaciones. Sobre la distribución de frecuencias se superpone una gráfica de la curva normal. En el capítulo 7 se le explicará la curva normal.

Minitab



Autoevaluación 3.8



Los años de servicio de una muestra de siete empleados en la oficina de quejas de State Farm Insurance en Cleveland, Ohio, son: 4, 2, 5, 4, 5, 2 y 6. ¿Cuál es la varianza de la muestra? Calcule la desviación estándar de la muestra.

Ejercicios

En los ejercicios 47-52, efectúe lo siguiente:

a) Calcule la varianza de la muestra;
b) Determine la desviación estándar de la muestra.

47. Considere los siguientes valores como una muestra: 7, 2, 6, 2 y 3.

48. Los siguientes cinco valores son una muestra: 11, 6, 10, 6 y 7.

49. Dave´s Automatic Door, referido en el ejercicio 37, instala puertas automáticas para cocheras. Sobre la base de una muestra, los siguientes son los tiempos, en minutos, que se requieren para instalar 10 puertas automáticas: 28, 32, 24, 46, 44, 40, 54, 38, 32 y 42.

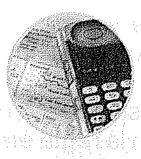
50. A la muestra de ocho compañías en la industria aeronáutica, referida en el ejercicio 36, se le aplicó una encuesta referente a su recuperación de inversión del año pasado. Los resultados son los siguientes: 10.6, 12.6, 14.8, 18.2, 12.0, 14.8, 12.2 y 15.6.

51. La Asociación de Propietarios de Moteles de Houston, Texas, llevó a cabo una encuesta relativa a las tarifas de motel entre semana en el área. Enseguida aparece la tarifa por cuarto para huéspedes de negocios en una muestra de 10 moteles.

\$101	\$97	\$103	\$110	\$78	\$87	\$101	\$80	\$106	\$88

52. Una organización de protección al consumidor se ocupa de las deudas con las tarjetas de crédito. Una encuesta entre 10 adultos jóvenes con una deuda con la tarjeta de crédito de más de \$2 000 mostró que éstos pagan en promedio un poco más de \$100 mensuales como abono a sus saldos. En la siguiente lista aparecen las sumas que cada adulto joven pagó el mes pasado.

\$110	\$126	\$103	\$93	\$99	\$113	\$87	\$101	\$109	\$100



Estadística en acción

Un promedio es un valor empleado para representar todos los datos. Sin embargo, a menudo no ofrece el panorama de los datos. Los inversionistas encaran con frecuencia con este problema cuando consideran dos inversiones en fondos mutualistas, como el Índice Vanguard 500 y los fondos GNMA. En agosto de 2003, la tasa de rendimiento anualizada de los fondos del Index 500 fue de –11.26% con una desviación estándar de 16.9. El fondo GNMA tuvo una tasa de rendimiento anualizada de 8.86% con una desviación estándar de 2.68. La desviación estándar muestra que la tasa de rendimiento del Índice 500 puede variar mucho. De hecho, las tasas de rendimiento anuales de los pasados 10 años variaron entre –22.15% a 37.45%. La desviación estándar del fondo GNMA es mucho menor. Sus tasas de rendimiento durante los pasados 10 años variaron de –0.95% a 11.22%. (www.vanguard.com)

Interpretación y usos de la desviación estándar

La desviación estándar normalmente se utiliza como medida para comparar la dispersión de dos o más conjuntos de observaciones. Por ejemplo, se calcula que la desviación estándar de las sumas quincenales invertidas en el plan de reparto de utilidades Dupree Saint Company es de \$7.51. Suponga que estos empleados se ubican en Georgia. Si la desviación estándar de un grupo de empleados en Texas es de \$10.47 y las medias son casi las mismas, esto indica que las sumas invertidas por los empleados de Georgia no se encuentran tan dispersas como las de los empleados en Texas (ya que \$7.51 < \$10.47). Como las sumas invertidas por los empleados de Georgia se acumulan más cerca de la media, la media para los empleados de Georgia es una medida más confiable que la media para el grupo de Texas.

Teorema de Chebyshev

Ya se ha insistido en el hecho de que una desviación estándar pequeña para un conjunto de valores, indica que estos valores se localizan cerca de la media. Por lo contrario, una desviación grande revela que las observaciones se encuentran muy dispersas con respecto a la media. El matemático ruso P. L. Chebyshev (1821-1894) estableció un teorema que nos permite determinar la mínima porción de valores que se encuentran a cierta cantidad de desviaciones estándares de la media. Por ejemplo, de acuerdo con el **teorema de Chebyshev**, por lo menos tres de cuatro valores, o 75%, deben encontrarse entre la media más dos desviaciones estándares y la media menos dos desviaciones estándares. Esta relación se cumple con independencia de la forma de la distribución. Además, por lo menos ocho de los nueve valores, 88.9%, se encontrarán más de tres desviaciones estándares y menos tres desviaciones estándares de la media. Por lo menos 24 de 25 valores, o 96%, se encontrará entre más y menos cinco desviaciones estándares de la media.

El teorema de Chebyshev establece lo siguiente:

TEOREMA DE CHEBYSHEV En cualquier conjunto de observaciones (muestra o población), la proporción de valores que se encuentran a k desviaciones estándares de la media es de por lo menos $1 - 1/k^2$, siendo k cualquier constante mayor que 1.

Ejemplo

La media aritmética de la suma quincenal que aportan los empleados de Dupree Saint para el plan de reparto de utilidades de la compañía es de $51.54 y la desviación estándar, de $7.51. ¿Por lo menos qué porcentaje de las aportaciones se encuentra en más 3.5 desviaciones estándares y menos 3.5 desviaciones de la media?

Solución

Alrededor de 92%, que se determina de la siguiente manera:

$$1-\frac{1}{k^2}=1-\frac{1}{(3.5)^2}=1-\frac{1}{12.25}=0.92$$

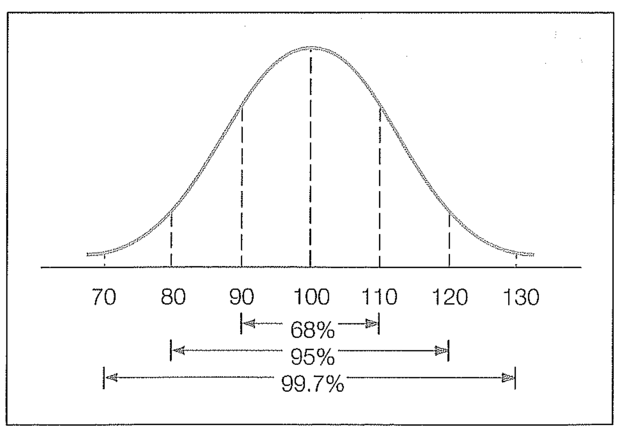
La regla empírica

La regla empírica sólo se aplica a distribuciones simétricas con forma de campana

El teorema de Chebyshev tiene que ver con cualquier conjunto de valores; es decir, que la distribución de valores puede tener cierta forma. Sin embargo, en cualquier distribución simétrica con forma de campana, como muestra la gráfica 3.7, es posible ser más precisos en la explicación de la dispersión en torno a la media. Estas relaciones que implican la desviación estándar y la media se encuentran descritas en la **regla empírica**, a veces denominada **regla normal**.

REGLA EMPÍRICA En cualquier distribución de frecuencias simétrica con forma de campana, aproximadamente 68% de las observaciones se encontrarán entre más y menos una desviación estándar de la media; cerca de 95% de las observaciones se encontrarán entre más y menos dos desviaciones estándares de la media y, de hecho todas (99.7%), estarán entre más y menos tres desviaciones estándares de la media.

Estas relaciones se representan en la gráfica 3.7 en el caso de una distribución con forma de campana con una media de 100 y una desviación estándar de 10.



GRÁFICA 3.7 Curva simétrica con forma de campana que muestra las relaciones entre la desviación estándar y las observaciones

Se ha observado que si una distribución es simétrica y tiene forma de campana, todas las observaciones se encuentran entre la media más y menos tres desviaciones estándares.

Por consiguiente, si $\bar{X} = 100$ y $s = 10$, todas las observaciones se encuentran entre $100 + 3(10)$ y $100 - 3(10)$, o 70 y 130. Por tanto, el rango es de 60, que se calcula restando $130 - 70$.

Por lo contrario, si sabe que el rango es de 60, puede aproximar la desviación estándar dividiendo el rango entre 6. En este caso: rango $\div 6 = 60 \div 6 = 10$, la desviación estándar.

Ejemplo

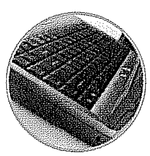
Una muestra de tarifas de renta de los departamentos University Park se asemeja a una distribución simétrica con forma de campana. La media de la muestra es de \$500; la desviación estándar de \$20. De acuerdo con la regla empírica conteste las siguientes preguntas:

1. ¿Entre qué dos cantidades se encuentra aproximadamente 68% de los gastos mensuales en alimentos?
2. ¿Entre qué dos cantidades se encuentra cerca de, 95% de los gastos mensuales en alimentos?
3. ¿Entre qué dos cantidades se encuentran casi todos los gastos mensuales en alimentos?

Solución

1. Cerca de 68% se encuentra entre \$480 y \$520, calculado de la siguiente manera: $\bar{X} \pm 1s = \$500 \pm 1(\$20)$.
2. Aproximadamente 95% se encuentra entre \$460 y \$540, calculado de la siguiente manera: $\bar{X} \pm 2s = \$500 \pm 2(\$20)$.
3. Casi todas (99.7%) se encuentran entre \$440 y \$560, calculado de la siguiente manera: $\bar{X} \pm 3s = \$500 \pm 3(\$20)$.

Autoevaluación 3.9



Pitney Pipe Company es uno de los fabricantes nacionales de tubos PVC. El departamento de control de calidad tomó una muestra de 600 tubos de 10 pies de longitud. A una distancia de 1 pie del extremo del tubo, se midió el diámetro externo. La media fue de 14.0 pulgadas y la desviación estándar de 0.1 pulgadas.

a) Si no conoce la forma de la distribución, ¿por lo menos qué porcentaje de las observaciones se encontrará entre 13.85 y 14.5 pulgadas?

b) Si supone que la distribución de los diámetros es simétrica y tiene forma de campana, ¿entre qué dos valores se encontrará aproximadamente 95% de las observaciones?

Ejercicios

53. De acuerdo con el teorema de Chebyshev, ¿por lo menos qué porcentaje de cualquier conjunto de observaciones se encontrará a 1.8 desviaciones estándares de la media?

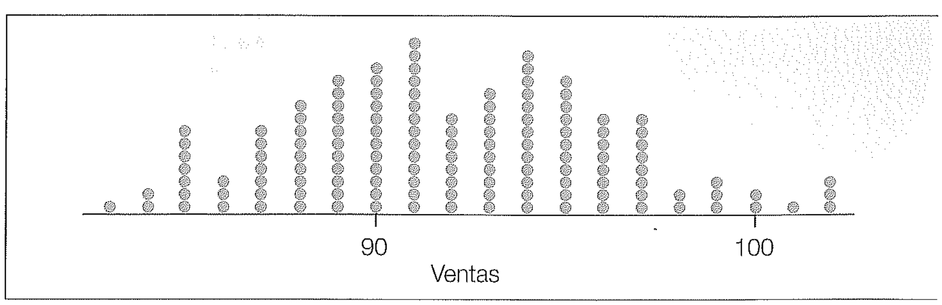
54. El ingreso medio de un grupo de observaciones de una muestra es de \$500; la desviación estándar es de \$40. De acuerdo con el teorema de Chebyshev, ¿por lo menos qué porcentaje de ingresos se encontrará entre \$400 y \$600?

55. La distribución de pesos de una muestra de 1 400 contenedores de carga es simétrica y tiene forma de campana. De acuerdo con la regla empírica, ¿qué porcentaje de pesos se encontrará entre:

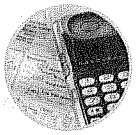
a) entre $\bar{X} - 2s$ y $\bar{X} + 2s$?;

b) ¿entre $\bar{X}$ y $\bar{X} + 2s$? ¿Debajo de $\bar{X} - 2s$?

56. La siguiente gráfica representa la distribución del número de refrescos tamaño gigante vendidos en un restaurante Wendy's los recientes 141 días. La cantidad promedio de refrescos vendidos por día es de 91.9 y la desviación estándar de 4.67.



Si utiliza la regla empírica, ¿entre qué dos valores de 68% de los días se encontrarán las ventas?



Estadística en acción

Derrek Lee, de los Osos de Chicago, ostentó el máximo promedio de bateo de 0.335 durante la temporada 2005. Tony Gwynn bateó 0.394 en la temporada 1994, en la que hubo pocos strikes, y Ted Williams bateó 0.406 en 1941. Nadie ha bateado arriba de 0.400 desde 1941. El promedio de bateo se ha mantenido constante alrededor de 0.260 por más de 100 años, pero la desviación estándar se redujo de 0.049 a 0.031. Esto indica que hay menos dispersión en el promedio de bateo de hoy y permite explicar la falta de bateadores que hayan alcanzado 0.400 recientemente.

La media y la desviación estándar de datos agrupados

En la mayoría de los casos las medidas de ubicación, como la media, y las medidas de dispersión, como la desviación estándar, se determinan utilizando valores individuales. Los paquetes de software de estadística facilitan el cálculo de estos valores, incluso en el caso de conjuntos grandes de datos. Sin embargo, algunas veces sólo se cuenta con la distribución de frecuencias y se desea calcular la media o la desviación estándar. En la siguiente discusión, se le mostrará cómo calcular la media y la desviación estándar a partir de datos organizados en una distribución de frecuencias. Hay que insistir en que una media o una desviación estándar de datos agrupados es una *estimación* de los valores reales correspondientes.

Media aritmética

Para aproximar la media aritmética de datos organizados en una distribución de frecuencia, comience suponiendo que las observaciones en cada clase se representan a través del *punto medio* de la clase. La media de una muestra de datos organizados en una distribución de frecuencias se calcula de la siguiente manera:

MEDIA ARITMÉTICA DE DATOS AGRUPADOS $$\bar{X} = \frac{\Sigma fM}{n} \qquad \textbf{[3.12]}$$

En esta fórmula:

$\bar{X}$ designa la media muestral;
M es el punto medio de cada clase;
f es la frecuencia en cada clase;
fM es la frecuencia en cada clase multiplicada por el punto medio de la clase;
ΣfM es la suma de estos productos;
n es el número total de frecuencias.

Ejemplo

Los cálculos de la media aritmética de datos agrupados en una distribución de frecuencias que aparecen enseguida se basan en los datos de Whitner Autoplex. Recuerde que en el capítulo 2, tabla 2.7, construyó una distribución de frecuencias de precios de venta de vehículos. La información se repite abajo. Determine el precio de venta medio aritmético de los vehículos.

Precio de venta (miles de dólares)	Frecuencia
15 a 18	8
18 a 21	23
21 a 24	17
24 a 27	18
27 a 30	8
30 a 33	4
33 a 36	2
Total	80

Solución

El precio de venta medio de los vehículos se calcula a partir de datos agrupados en una distribución de frecuencias. Para calcular la media, suponga que el punto medio de cada clase es representativo de los valores de datos en dicha clase. Recuerde que el punto medio de una clase se encuentra a la mitad de los límites de clase superior e inferior. Para determinar el punto medio de una clase en particular, sume los límites de clase superior e inferior y divida entre 2. Por consiguiente, el punto medio de la primera clase es $16.5, que se calcula con la operación ($15 + $18)/2. Asuma que el valor de $16.5 es representativo de los ocho valores en dicha clase. En otras palabras, se asume que la suma de los ochos valores en esta clase es de $132, que se calcula por medio del producto 8($16.5). Continúe con el proceso de multiplicación del punto medio de clase por la frecuencia de clase de cada clase y enseguida sume estos productos. Los resultados se resumen en la tabla 3.1.

TABLA 3.1 Precio de 80 nuevos vehículos vendidos el mes pasado en el lote de Whitner Autoplex

Precio de venta (miles de dólares)	Frecuencia (f)	Punto medio (M)	fM
15 a 18	8	$16.5	$ 132.0
18 a 21	23	19.5	448.5
21 a 24	17	22.5	382.5
24 a 27	18	25.5	459.0
27 a 30	8	28.5	228.0
30 a 33	4	31.5	126.0
33 a 36	2	34.5	69.0
Total	80		$1 845.0

Al despejar la media aritmética de la fórmula 3.12 se obtiene:

$$\bar{X} = \frac{\Sigma fM}{n} = \frac{\$1\,845}{80} = \$23.1\,(\text{miles})$$

Así, se concluye que el precio de venta medio de los vehículos es de aproximadamente $23 100.

Desviación estándar

Para calcular la desviación estándar de datos agrupados en una distribución de frecuencias, necesita ajustar ligeramente la fórmula 3.11. Pondere cada una de las diferencias cuadradas por el número de frecuencias en cada clase. La fórmula es:

DESVIACIÓN ESTÁNDAR, DATOS AGRUPADOS

$$s = \sqrt{\frac{\Sigma f(M - \bar{X})^2}{n-1}} \qquad \textbf{[3.13]}$$

en la que:

s es el símbolo de la desviación estándar de la muestra;
M es el punto medio de la clase;
f es la frecuencia de clase;
n es el número de observaciones en la muestra;
$\bar{X}$ designa la media muestral.

Ejemplo

Consulte la distribución de frecuencias de los datos de Whitner Autoplex que aparecen en la tabla 3.1. Calcule la desviación estándar de los precios de venta de los vehículos.

Solución

De acuerdo con la misma técnica empleada anteriormente para calcular la media de los datos agrupados en una distribución de frecuencias, *f* es la frecuencia de clase, *M* es el punto medio de clase y *n* es el número de observaciones.

Precio de venta (miles de dólares)	Frecuencia (f)	Punto medio (M)	$(M - \bar{X})$	$(M - \bar{X})^2$	$f(M - \bar{X})^2$
15 a 18	8	16.5	−6.6	43.56	348.48
18 a 21	23	19.5	−3.6	12.96	298.08
21 a 24	17	22.5	−0.6	0.36	6.12
24 a 27	18	25.5	2.4	5.76	103.68
27 a 30	8	28.5	5.4	29.16	233.28
30 a 33	4	31.5	8.4	70.56	282.24
33 a 36	2	34.5	11.4	129.96	259.92
	80				1 531.80

Para determinar la desviación estándar:

Paso 1: Reste la media del punto medio de clase. Es decir, encuentre $(M - \bar{X})$. Para la primera clase $(16.5 - 23.1 = -6.6)$; para la segunda clase $(19.5 - 23.1 = -3.6)$ y así en lo sucesivo.

Paso 2: Eleve al cuadrado la diferencia entre el punto medio de clase y la media. En el caso de la primera clase sería $(16.5 - 23.1)^2 = (-6.6)^2 = 43.56$; en el caso de la segunda clase $(19.5 - 23.1)^2 = (-3.6)^2 = 12.96$ y así en lo sucesivo.

Paso 3: Multiplique la diferencia al cuadrado entre el punto medio de clase y la media por la frecuencia de clase. Para la primera clase el valor es $8(16.5 - 23.1)^2 = 348.48$; para la segunda $23(19.5 - 23.1)^2 = 298.08$ y así sucesivamente.

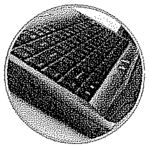
Paso 4: Sume $f(M - \bar{X})^2$. El total es 1 531.8.

Para determinar la desviación estándar, sustituya estos valores en la fórmula 3.13.

$$s = \sqrt{\frac{\Sigma f(M - \bar{X})^2}{n-1}} = \sqrt{\frac{1531.8}{80-1}} 4.403$$

La media y la desviación estándar calculadas a partir de datos agrupados en una distribución de frecuencias, por lo general se encuentran cerca de los valores calculados a partir de los datos en bruto. Los datos agrupados originan la pérdida de alguna información. En el caso del problema del precio de venta de los vehículos, el precio medio de venta que aparece en la hoja de Excel de la página 66 es de \$23 218 y la desviación estándar de \$4 354. Los valores respectivos calculados a partir de datos agrupados en una distribución de frecuencias son \$23 100 y \$4 403. La diferencia en las medias es de \$118 o aproximadamente 0.51%. Las desviaciones estándares difieren por \$49 o 1.1%. Sobre la base de la diferencia porcentual, las aproximaciones se acercan mucho a los valores reales.

Autoevaluación 3.10



Los ingresos netos de una muestra de grandes importadores de antigüedades se organizaron en la siguiente tabla:

Ingreso neto (millones de dólares)	Número de importadores
2 a 6	1
6 a 10	4
10 a 14	10
14 a 18	3
18 a 22	2

a) ¿Qué nombre recibe la tabla?
b) Sobre la base de la distribución, ¿cuál es el cálculo aproximado del ingreso neto medio aritmético?
c) Con base en la distribución, ¿cuál es el cálculo aproximado de la desviación estándar?

Ejercicios

57. Cuando calcula la media de una distribución de frecuencia, ¿por qué hace referencia a ésta como una media *aproximada?*

58. Determine la media y la desviación estándar de la siguiente distribución de frecuencias.

Clase	Frecuencia
0 a 5	2
5 a 10	7
10 a 15	12
15 a 20	6
20 a 25	3

59. Determine la media y la desviación estándar de la siguiente distribución de frecuencias.

Clase	Frecuencia
20 a 30	7
30 a 40	12
40 a 50	21
50 a 60	18
60 a 70	12

60. SCCoast, un proveedor de Internet en el sureste de Estados Unidos, elaboró una distribución de frecuencias sobre la edad de los usuarios de Internet. Determine la media y la desviación estándar.

Edad (años)	Frecuencia
10 a 20	3
20 a 30	7
30 a 40	18
40 a 50	20
50 a 60	12

61. El IRS (Internal Revenue Service) estaba interesado en el número de formas fiscales individuales que preparan las pequeñas empresas de contabilidad. El IRS tomó una muestra aleatoria de 50 empresas de contabilidad pública con 10 o más empleados en la zona de Dallas-Fort Worth. La siguiente tabla de frecuencias muestra los resultados del estudio. Calcule la media y la desviación estándar.

Número de clientes	Frecuencia
20 a 30	1
30 a 40	15
40 a 50	22
50 a 60	8
60 a 70	4

62. Los gastos en publicidad constituyen un elemento significativo del costo de los artículos vendidos. Enseguida aparece una distribución de frecuencias que muestra los gastos en publicidad de 60 compañías fabricantes ubicadas en el suroeste de Estados Unidos. Calcule la media y la desviación estándar de los gastos de publicidad.

Gastos en publicidad (millones de dólares)	Número de compañías
25 a 35	5
35 a 45	10
45 a 55	21
55 a 65	16
65 a 75	8
Total	60

Ética e informe de resultados

En el capítulo 1 se analizó la manera de informar resultados estadísticos con ética e imparcialidad. Aunque está aprendiendo a organizar, resumir e interpretar datos empleando la estadística, también es importante que comprenda la estadística con el fin de que se convierta en un consumidor de información inteligente.

En este capítulo, aprendió la forma de calcular estadísticas descriptivas de naturaleza numérica. Específicamente la manera de calcular e interpretar medidas de ubicación para un conjunto de datos: la media, la mediana y la moda. También ha estudiado las ventajas y desventajas de cada estadístico. Por ejemplo, si un agente de bienes raíces le dice a un cliente que la casa promedio de determinada parcela se vendió en $150 000, supondrá que $150 000 es un precio de venta representativo de todas las casas. Pero si el cliente pregunta, además, cuál es la mediana del precio de venta y resulta ser $60 000. ¿Por qué informó el agente solamente el precio promedio? Esta información es de suma importancia para que una persona tome una decisión cuando compra una casa. Conocer las ventajas y desventajas de la media, la mediana y la moda es importante al dar un informe estadístico y cuando se emplea información estadística para tomar decisiones.

También aprendió a calcular medidas de dispersión: el rango, la desviación media y la desviación estándar. Cada uno de estos estadísticos también tiene ventajas y desventajas. Recuerde que el rango proporciona información sobre la dispersión total de una distribución. Sin embargo, no proporciona información sobre la forma en que se acumulan los datos o se concentran en torno al centro de la distribución.

Conforme aprenda más estadística, necesitará recordar que cuando emplea estadísticas debe mantener un punto de vista independiente y con principios. Cualquier informe estadístico requiere la comunicación honesta y objetiva de los resultados.

Resumen del capítulo

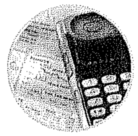
I. Una medida de ubicación es un valor que sirve para describir el centro de un conjunto de datos.

  A. La media aritmética es la medida de ubicación que más se informa.

    1. Se calcula sumando los valores de las observaciones y dividiendo entre el número total de observaciones.

      a) La fórmula para una media poblacional de datos no agrupados o en bruto es:

$$\mu = \frac{\Sigma X}{N} \qquad \textbf{[3.1]}$$



Estadística en acción

La mayoría de las universidades informan el *tamaño promedio de los grupos*. Esta información puede inducir a error, ya que el tamaño promedio de los grupos se determina de diversas formas. Si calcula la cantidad de estudiantes en cada clase en cierta universidad, el resultado es la cantidad promedio de estudiantes por clase. Si recaba una lista de tamaños de grupos y calcula el tamaño de grupo promedio, podría hallar que la media es muy diferente. Una escuela descubrió que el promedio de estudiantes en cada una de sus 747 clases era de 40. Pero cuando calculó la media a partir de una lista de tamaños de grupo, ésta resultó ser de 147. ¿Por qué la discrepancia? Hay menos estudiantes en los grupos pequeños y una gran cantidad de estudiantes en los grupos grandes, lo cual tiene el efecto de incrementar el tamaño promedio de los grupos cuando se calcula de esta manera. Una universidad podría reducir su tamaño promedio de grupo reduciendo el número de estudiantes en cada grupo. Esto significa eliminar las cátedras en las que hay muchos estudiantes de primer grado.

b) La fórmula para la media de una muestra es:

$$\bar{X} = \frac{\Sigma X}{n} \qquad \textbf{[3.2]}$$

c) La fórmula para la media muestral en una distribución de frecuencias es.

$$\bar{X} = \frac{\Sigma fM}{n} \qquad \textbf{[3.12]}$$

2. Las características principales de la media aritmética son las siguientes:
a) Por lo menos se requiere la escala de medición de intervalo.
b) Todos los valores de los datos se incluyen en el cálculo.
c) Un conjunto de datos sólo posee una media. Es decir que ésta es única.
d) La suma de las desviaciones de la media es igual a 0.

B. La media ponderada se encuentra multiplicando cada observación por su correspondiente ponderación.
1. La fórmula para determinar la media ponderada es:

$$\bar{X}_w = \frac{W_1X_1 + W_2X_2 + W_3X_3 + \cdots + W_nX_n}{W_1 + W_2 + W_3 + \cdots + W_n} \qquad \textbf{[3.3]}$$

2. Éste es un caso especial de la media aritmética.

C. La mediana es el valor que se encuentra en medio de un conjunto de datos ordenados.
1. Para determinar la mediana, se ordenan las observaciones de menor a mayor y se identifica el valor intermedio.
2. Las principales características de la mediana son las siguientes:
a) Se requiere por lo menos la escala ordinal de medición.
b) No influyen sobre ésta valores extremos.
c) Cincuenta por ciento de las observaciones son más grandes que la mediana.
d) Ésta es única para un conjunto de datos.

D. La moda es el valor que se presenta con mayor frecuencia en un conjunto de datos.
1. La moda se determina en el caso de datos de nivel nominal.
2. Un conjunto de datos puede tener más de una moda.

E. La media geométrica es la enésima raíz del producto de n valores positivos.
1. La fórmula de la media geométrica es la siguiente:

$$GM = \sqrt[n]{(X_1)(X_2)(X_3)\cdots(X_n)} \qquad \textbf{[3.4]}$$

2. La media geométrica también se emplea para determinar la razón de cambio de un periodo a otro.

$$GM = \sqrt[n]{\frac{\text{Valor al final del periodo}}{\text{Valor al principio del periodo}}} \qquad \textbf{[3.5]}$$

3. La media geométrica siempre es igual o menor que la media aritmética.

II. La dispersión es la variación o propagación en un conjunto de datos.

A. El rango es la diferencia entre el valor máximo y el mínimo en un conjunto de datos.
1. La fórmula del rango es la siguiente:

$$\text{Rango} - \text{Valor más alto} - \text{Valor más bajo} \qquad \textbf{[3.6]}$$

2. Las principales características del rango son:
a) Sólo dos valores se emplean en su cálculo.
b) Recibe la influencia de los valores extremos.
c) Es fácil de calcular y definir.

B. La desviación absoluta media es la suma de los valores absolutos de las desviaciones de la media, dividida entre el número de observaciones.
1. La fórmula para calcular la desviación absoluta media es:

$$MD = \frac{\Sigma|X - \bar{X}|}{n} \qquad \textbf{[3.7]}$$

2. Las principales características de la desviación absoluta media son las siguientes:
a) No influyen excesivamente sobre ella valores grandes o pequeños.
b) Todas las observaciones se emplean en el cálculo.
c) Los valores absolutos son de alguna forma difíciles de manejar.

C. La varianza es la media de las desviaciones al cuadrado de la media aritmética.

1. La fórmula de la varianza de la población es la siguiente:

$$\sigma^2 = \frac{\Sigma(X-\mu)^2}{N} \qquad \textbf{[3.8]}$$

2. La fórmula de la varianza de la muestra es la siguiente:

$$s^2 = \frac{\Sigma(X-\bar{X})^2}{n-1} \qquad \textbf{[3.10]}$$

3. Las principales características de la varianza son:
   - ***a)*** Todas las observaciones se utilizan en el cálculo.
   - ***b)*** No influyen excesivamente sobre ella observaciones extremas.
   - **c)** Resulta de alguna manera difícil trabajar con las unidades; éstas son las unidades originales elevadas al cuadrado.

D. La desviación estándar es la raíz cuadrada de la varianza.

1. Las principales características de la desviación estándar son:
   - ***a)*** Se expresa en las mismas unidades de los datos originales.
   - ***b)*** Es la raíz cuadrada de la distancia promedio al cuadrado de la media.
   - **c)** No puede ser negativa.
   - ***d)*** Es la medida de dispersión que se informa con más frecuencia.
2. La fórmula de la desviación estándar de la muestra es:

$$s = \sqrt{\frac{\Sigma(X-\bar{X})^2}{n-1}} \qquad \textbf{[3.11]}$$

3. La fórmula de la desviación estándar para datos agrupados es:

$$s = \sqrt{\frac{\Sigma f(M-\bar{X})^2}{n-1}} \qquad \textbf{[3.13]}$$

III. Interpretó la desviación estándar empleando dos medidas.

A. El teorema de Chebyshev establece que independientemente de la forma de la distribución, por lo menos $1 - 1/k^2$ de las observaciones se encontrarán a k desviaciones estándares de la media, siendo k mayor que 1.

B. La regla empírica afirma que en el caso de una distribución en forma de campana, aproximadamente 68% de los valores se encontrarán a una desviación estándar de la media; 95%, a dos y casi todas, a tres.

Clave de pronunciación

SÍMBOLO	SIGNIFICADO	PRONUNCIACIÓN
μ	Media de población	*Mu*
Σ	Operación de suma	*Sigma*
ΣX	Suma de un grupo de valores	*Sigma X*
$\bar{X}$	Media de la muestra	*X barra*
$\bar{X}_w$	Media ponderada	*X barra subíndice w*
GM	Media geométrica	*G M*
ΣfM	Suma del producto de las frecuencias y los puntos medios de clase	*Sigma f M*
σ^2	Varianza de la población	*Sigma al cuadrado*
σ	Desviación estándar de la población	*Sigma*

Ejercicios del capítulo

63. La empresa de contabilidad Crawford and Associates posee cinco socios. El día de ayer los socios atendieron a seis, cuatro, siete y cinco clientes, respectivamente.

- ***a)*** Calcule el número medio y el número mediano de clientes que cada socio atendió.
- ***b)*** ¿Es la media una muestral o una poblacional?
- **c)** Verifique que $\Sigma(X-\mu) = 0$.

64. Owens Orchards vende manzanas por peso en bolsas grandes. Una muestra de siete bolsas contenía las siguientes cantidades de manzanas: 23, 19, 26, 17, 21, 24 y 22.

a) Calcule la cantidad media y la cantidad mediana de manzanas que hay en una bolsa.

b) Verifique que $\Sigma(X - \bar{X}) = 0$.

65. Una muestra de familias que ha contratado los servicios de la United Bell Phone Company reveló el siguiente número de llamadas recibidas por familia la semana pasada. Determine el número medio y la mediana de llamadas recibidas.

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

66. La Citizens Banking Company estudia la cantidad de veces que utiliza al día el cajero automático ubicado en uno de los supermercados de Loblaws, sobre Market Steet. Enseguida figuran las cantidades de ocasiones que se utilizó la máquina al día durante los pasados 30 días. Determine la cantidad media de veces que se utilizó la máquina al día.

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

67. El gobierno canadiense desea conocer la edad relativa de su fuerza laboral. Conforme la generación de *baby boomers* envejece, el gobierno se interesa en la disponibilidad de trabajadores jóvenes calificados. Con el fin de informarse, el gobierno realiza una encuesta en varias industrias sobre las edades de los empleados. La siguiente tabla contiene la edad media y mediana para dos industrias, comunicaciones y comercio minorista, tomando en cuenta seis diferentes tipos de trabajo.

	Comunicación y otras empresas		**Comercio minorista y servicios al consumidor**	
	Media	Mediana	Media	Mediana
Directores	42.6	43	38.6	38
Profesionales	40.8	40	40.0	39
Técnica/Oficios	41.4	42	37.1	37
Marketing/Ventas	NA	NA	33.7	31
Oficinistas/Administrativos	40.8	41	38.0	38
Trabajadores de la producción	37.2	40	32.0	24

Comente sobre la distribución de edades. ¿Qué industria parece tener trabajadores de más edad? ¿Cuál tiene trabajadores más jóvenes? ¿Qué tipos de trabajo muestran la mayor diferencia entre la edad media y la mediana en cada industria?

68. Trudy Green trabaja para la True-Green Lawn Company. Su trabajo consiste en buscar por teléfono negocios de mantenimiento de césped. Enseguida aparece una lista de la cantidad de citas por hora que hizo durante las últimas 25 horas de llamadas. ¿Cuál es la media aritmética de citas que hace por hora? ¿Cuál es la cantidad mediana de citas que hace por hora? Redacte un breve informe que resuma sus conclusiones.

9	5	2	6	5	6	4	4	7	2	3	6	3
4	4	7	8	4	4	5	5	4	8	3	3	

69. La Split-A-Rail Fence Company vende tres tipos de cerca a propietarios de casa en los suburbios de Seattle, Washington. Las cercas grado A tienen un costo de $5.00 el pie de instalación. Las cercas grado B tienen un costo de $6.50 el pie de instalación y las grado C, las de alta calidad, tienen un costo de $8.00 el pie de instalación. Ayer, Split-A-Rail instaló 270 pies de cerca grado A, 300 pies de cerca grado B y 100 pies de cerca grado C. ¿Cuál fue el costo medio por pie de cerca instalada?

70. Rolland Poust es un estudiante de primer grado de la Facultad de Administración del Scandia Tech. El semestre anterior tomó dos cursos de estadística y contabilidad de 3 horas cada uno y obtuvo una A en ambos. Obtuvo B en un curso de historia de cinco horas y B en un curso de historia del jazz de dos horas. Además tomó un curso de una hora que tenía que ver con las reglas de básquetbol con el fin de obtener su licencia para arbitrar partidos de básquetbol de escuela secundaria. Obtuvo una A en este curso. ¿Cuál fue su promedio semestral? Suponga que le dan 4 puntos por una A; 3 por una B y así sucesivamente. ¿Qué medida de ubicación calculó?

71. La siguiente tabla muestra el porcentaje de fuerza laboral desempleada y el tamaño de la fuerza laboral en tres condados del noroeste de Ohio. Jon Elsas es director regional de desarrollo económico. Debe presentar un informe a varias compañías que piensan ubicarse en el noroeste de Ohio. ¿Cuál sería un índice de desempleo adecuado para toda la región?

Condado	Porcentaje de desempleo	Tamaño de la fuerza laboral
Wood	4.5	15 300
Ottawa	3.0	10 400
Lucas	10.2	150 600

72. La American Automobile Association verifica los precios de la gasolina antes de varios fines de semana festivos. La siguiente lista incluye los precios de autoservicio de una muestra de 15 gasolinerías de menudeo durante el fin de semana del día del trabajo de 2005 en el área de Detroit, Michigan.

3.44	3.42	3.35	3.39	3.49	3.49	3.41	3.46
3.41	3.49	3.45	3.48	3.39	3.46	3.44	

a) ¿Cuál es la media aritmética del precio de venta?
b) ¿Cuál es la mediana del precio de venta?
c) ¿Cuál es el precio de venta modal?

73. El área metropolitana de Los Ángeles-Long Beach, California, es el área que se espera que muestre el mayor incremento en el número de trabajos de 1989 a 2010. Se espera que el número de trabajos se incremente de 5 164 900 a 6 286 800. ¿Cuál es la media geométrica de la tasa de incremento anual esperada?

74. Un artículo reciente sugirió que, si en la actualidad usted gana $25 000 anuales y la tasa de inflación continúa siendo de 3% anual, usted necesitará ganar $33 598 en 10 años para tener el mismo poder adquisitivo. ¿Qué necesitaría hacer para percibir $44 771 si la tasa de inflación se elevara a 6%? Confirme si estas afirmaciones son exactas determinando la tasa media geométrica de incremento.

75. Las edades de una muestra que se tomó de turistas canadienses que vuelan de Toronto a Hong Kong fueron las siguientes: 32, 21, 60, 47, 54, 17, 72, 55, 33 y 41.
a) Calcule el rango.
b) Estime la desviación media.
c) Calcule la desviación estándar.

76. Los pesos (en libras) de una muestra de cinco cajas enviadas por UPS son: 12, 6, 7, 3 y 10.
a) Calcule el rango.
b) Aproxime la desviación media.
c) Calcule la desviación estándar.

77. Un estado del sur de Estados Unidos cuenta con siete universidades estatales en su sistema. Los números en volumen (en miles) que guardan en sus bibliotecas son: 83, 510, 33, 256, 401, 47 y 23.
a) ¿Es una muestra o una población?
b) Calcule la desviación estándar.

78. Los temas de salud representan una preocupación para gerentes, especialmente cuando éstos evalúan el costo del seguro médico. Una encuesta reciente de 150 ejecutivos de Elvers Industries, una importante empresa financiera y de seguros, ubicada en el suroeste de Estados Unidos, informó la cantidad de libras de sobrepeso de los ejecutivos. Calcule la media y la desviación estándar.

Libras de sobrepeso	Frecuencia
0 a 6	14
6 a 12	42
12 a 18	58
18 a 24	28
24 a 30	8

79. El programa espacial Apolo duró de 1967 hasta 1972 e incluyó 13 misiones. Las misiones tuvieron una duración de 7 a 301 horas. Enseguida aparece la duración de cada vuelo.

9	195	241	301	216	260	7	244	192	147
10	295	142							

a) Explique la razón por la que los tiempos de vuelo constituyen una población.
b) Calcule la media y la mediana de los tiempos de vuelo.
c) Estime el rango y la desviación estándar de los tiempos de vuelo.

80. Creek Ratz es un restaurante muy popular localizado en la costa del norte de Florida, sirve una variedad de alimentos con carne de res y mariscos. Durante la temporada de vacaciones de verano, no se aceptan reservaciones. La gerencia del restaurante está interesada en conocer el tiempo que un cliente tiene que esperar antes de pasar a la mesa. A continuación aparece la lista de tiempos de espera, en minutos, para las 25 mesas que se ocuparon la noche del sábado pasado.

28	39	23	67	37	28	56	40	28	50
51	45	44	65	61	27	24	61	34	44
64	25	24	27	29					

a) Explique la razón por la que los tiempos constituyen una población.
b) Calcule la media y la mediana de los tiempos de espera.
c) Estime el rango y la desviación estándar de los tiempos de espera.

81. El gerente de la tienda Wal-Mart de la localidad estudia la cantidad de artículos que compran los consumidores en el horario de la tarde. A continuación aparece la cantidad de artículos de una muestra de 30 consumidores.

15	8	6	9	9	4	18	10	10	12
12	4	7	8	12	10	10	11	9	13
5	6	11	14	5	6	6	5	13	5

a) Calcule la media y la mediana de la cantidad de artículos.
b) Estime el rango y la desviación estándar de la cantidad de artículos.
c) Organice la cantidad de artículos en una distribución de frecuencias. Quizá desee repasar las instrucciones del capítulo 2 para establecer el intervalo de clase y el número de clases.
d) Calcule la media y la desviación estándar de los datos organizados en una distribución de frecuencias. Compare estos valores con los que calculó en el inciso ***a)*** ¿Por qué son diferentes?

82. La siguiente distribución de frecuencias contiene los costos de electricidad de una muestra de 50 departamentos de dos recámaras en Albuquerque, Nuevo México, durante el mes de mayo del año pasado.

Costos de electricidad	Frecuencia
$ 80 a $100	3
100 a 120	8
120 a 140	12
140 a 160	16
160 a 180	7
180 a 200	4
Total	50

a) Calcule el costo medio.
b) Aproxime la desviación estándar.
c) Utilice la regla empírica para calcular la fracción de costos que se encuentra a dos desviaciones estándares de la media. ¿Cuáles son estos límites?

83. Bidwell Electronics, Inc., recién tomó una muestra de empleados para determinar la distancia a la que viven de las oficinas centrales de la empresa. Los resultados aparecen a continuación. Calcule la media y la desviación estándar.

Distancia (miles)	Frecuencia	M
0 a 5	4	2.5
5 a 10	15	7.5
10 a 15	27	12.5
15 a 20	18	17.5
20 a 25	6	22.5

ejercicios.com

84. El estado de Indiana y la Escuela de Administración Kelley de la Universidad de Indiana ofrecen vínculos para diversas fuentes de datos. Diríjase a www.stats.indiana.edu; enseguida, bajo el encabezado de indicadores sociales y económicos, seleccione *Birth/Death/Marriage;* bajo comparaciones de estados, seleccione *Anual Birth Data;* para Geography Type, seleccione *U.S. and 50 States;* para Specific Geography, seleccione *all states* y, finalmente, seleccione *Get Data.* La información se puede presentar en un formato de Excel. Suponga que se encuentra interesado en la cantidad típica de nacimientos por estado. Calcule la media, la mediana y la desviación estándar del *número de nacimientos por estado* y del *número de nacimientos por cada 1 000 habitantes por estado* para el último año disponible. Usted podría bajar esta información en un paquete de software para llevar a cabo los cálculos. ¿Qué medida de ubicación es la más representativa? ¿Qué conjunto de datos recomendaría utilizar: el *número de nacimientos por estado* o el *número de nacimientos por cada 1 000 habitantes?* ¿Por qué? Asuma que se encuentra interesado en las tasas de nacimiento de los 50 estados y de Washington, D. C. Calcule la media, la mediana y la desviación estándar. Redacte un breve informe que resuma los datos.

85. Existen muchos sitios Web de finanzas que proporcionan información sobre acciones por industria. Por ejemplo, diríjase a http://finance.yahoo.com y seleccione **Stock Research**; bajo **Analyst Research**, seleccione **Sector/Industry Analysis**. Aquí hay muchas opciones disponibles, como **Healthcare.** Ahora se abre otra lista de opciones; seleccione una, como **Drug Manufacturers-Major**. Aparecerá una lista de compañías en dicha industria. Elija una de las variables que aparecen, como la razón del precio respecto de las ganancias, que se encuentra representada por **P/E**. Esta variable es la razón del precio de venta de una acción de las acciones ordinarias de la compañía respecto de las ganancias por acción de las acciones ordinarias. Descargue esta información en Excel y determine la media, la mediana y la desviación estándar. Regrese a **Sector/Industry Analysis** y seleccione otro sector o industria. Tal vez desee seleccionar **Utilities** y, enseguida, **Gas Utilities**. Aparecerá una lista de compañías. Seleccione la misma variable que antes. Descargue la información en Excel y determine la media, la mediana y la desviación estándar para esta industria. Compare la información de los dos sectores. Redacte un breve informe que resuma sus conclusiones. ¿Son diferentes las medias? ¿Se presenta mayor variabilidad en una industria que en la otra?

86. Uno de los promedios más famosos, el Promedio Industrial Dow Jones (DJIA), no es realmente un promedio. A continuación aparece una lista de 30 compañías cuyos precios accionarios conforman el DJIA, su símbolo, su peso actual y el valor de cierre en agosto de 2005. Utilice un paquete de software para determinar la media de las 30 acciones. El DJIA es de 10 451. ¿Es el valor que usted encontró para el promedio de las 30 acciones?

Compañía	Símbolo	Precio	Compañía	Símbolo	Precio
Alcoa Inc.	AA	27.29	Johnson & Johnson	JNJ	61.94
Amer. Intl. Group	AIG	59.27	JP Morgan Chase	JPM	33.65
American Express	AXP	55.01	Coca-Cola Co.	KO	43.57
Boeing Co.	BA	66.31	McDonald's Corp.	MCD	33.48
Citigroup Inc.	C	43.10	3M Co.	MMM	70.99
Caterpillar Inc.	CAT	53.49	Altria Group Inc.	MO	69.48
Disney (Walt) Co.	DIS	25.33	Merck & Co.	MRK	27.66
DuPont (El)	DD	39.74	Microsoft Corp.	MSFT	26.97
General Electric	GE	33.38	Pfizer Inc.	PFE	24.89
General Motors	GM	34.14	Procter & Gamble	PG	54.96
Home Depot Inc.	HD	39.81	SBC Communication	SBC	23.71
Honeywell Intl.	HON	38.02	United Tech Corp.	UTX	50.29
Hewlett-Packard	HPQ	27.01	Verizon Communications	VZ	32.60
IBM	IBM	80.38	Wal-Mart Stores	WMT	45.70
Intel Corp.	INTC	25.41	Exxon Mobil Corp.	XOM	58.41

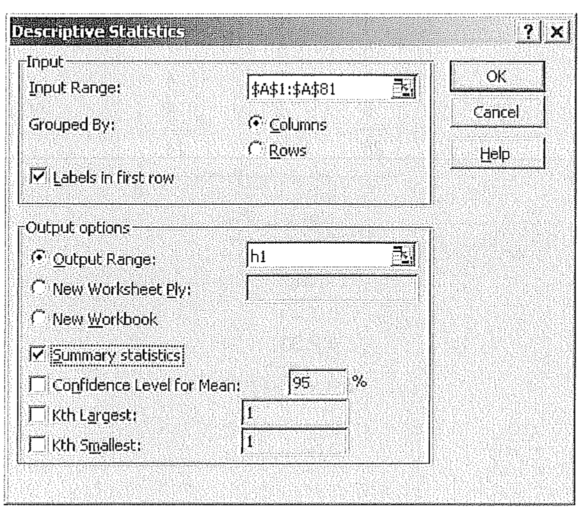
Puede leer sobre la historia de DJIA, diríjase a http://www.djindexes.com, haciendo clic en **About the Dow**. Aquí se explica la razón por la que no es un promedio. Hay muchos otros sitios que puede visitar para verificar el valor actual del DJIA: http://money.cnn.com, http://www.foxnews.com y http://www.usatoday.com son tres de las muchas fuentes. Para obtener una lista de las acciones reales que constituyen el promedio, diríjase a http://www.bloomberg.com. En la barra de herramientas, haga clic en **Market Data**; enseguida, bajando por la izquierda de la pantalla, seleccione **Stocks** y luego **Dow**. Aparecerá una lista de precios de venta actuales de 30 acciones que conforman el DJIA.

Ejercicios de la base de datos

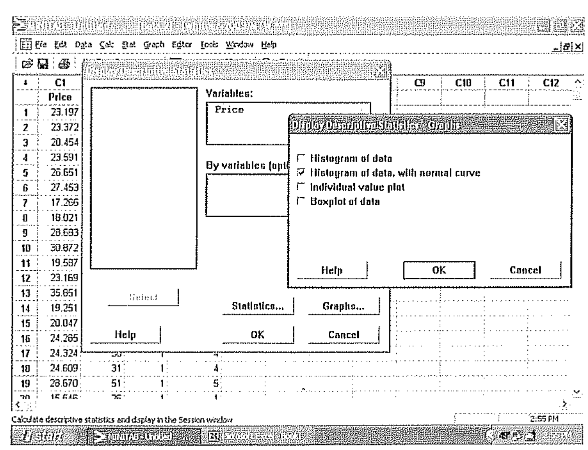
87. Consulte los datos Real Estate, que contienen información sobre casas vendidas en el área de Denver, Colorado, el año pasado.
    a) Seleccione la variable que se refiere al precio de venta.
        1. Determine la media, la mediana y la desviación estándar.
        2. Redacte un breve informe sobre la distribución de los precios de venta.
    b) Seleccione la variable que se refiere al área de la casa en pies cuadrados.
        1. Determine la media, la mediana y la desviación estándar.
        2. Redacte un breve informe sobre la distribución del área de las casas.
88. Consulte los datos Baseball 2005, que incluyen información sobre los 20 equipos de la liga mayor para la temporada 2005.
    a) Seleccione la variable que se refiere a los salarios de los equipos y calcule la media, la mediana y la desviación estándar.
    b) Seleccione la variable que se refiere a la fecha en que se construyó el estadio. (Sugerencia: reste el año en que se construyó el estadio del año actual para determinar la edad del estadio y trabaje con dicha variable.) Calcule la media, la mediana y la desviación estándar.
    c) Seleccione la variable que se refiere al cupo del estadio. Determine la media, la mediana y la desviación estándar.
89. Consulte los datos CIA, que proporcionan información demográfica y económica de 46 países.
    a) Seleccione la variable que se refiere a la expectativa de vida.
        1. Determine la media, la mediana y la desviación estándar.
        2. Redacte un breve resumen sobre la distribución de la expectativa de vida.
    b) Seleccione la variable *GDP/cap*.
        1. Calcule la media, la mediana y la desviación estándar.
        2. Redacte un breve resumen de la distribución *GDP/cap*.

Comandos de software

1. Los comandos de Excel de estadística descriptiva de la página 66 son los siguientes:
    a) Del CD recupere el archivo de datos Whitner, llamado **Whitner-2005**.
    b) De la barra de menú, seleccione **Tools** y, enseguida, **Data Analysis**. Seleccione **Descriptive Statistics** y, enseguida, haga clic en **OK**.
    c) Para **Input Range**, escriba *A1:A81*, indique que los datos se agrupan por columna y que las etiquetas se encuentran en la primera fila. Haga clic en **Output Range**, indique que la salida debe incluirse en *H1* (o en cualquier lugar que desee), haga clic en **Summary statistics** y, enseguida, en **OK**.



    d) Después de que obtenga los resultados, verifique dos veces la cuenta en la salida para cerciorarse de que contiene la cantidad correcta de elementos.
2. Los comandos de MINITAB para el resumen descriptivo de la página 80 son los siguientes:
    a) Del CD recupere los datos Whitner, llamados **Whitner 2005**.
    b) Seleccione **Stat, Basic Statistics** y, enseguida, **Display Descriptive Statistics**. En el cuadro de diálogo seleccione *Price* como variable y, enseguida, haga clic en **Graphs** en la esquina inferior derecha. Dentro del nuevo cuadro de diálogo seleccione **Histogram of data, with normal curve** y haga clic en **OK**. Haga clic en **OK** en el siguiente cuadro de diálogo.



Capítulo 3 Respuestas a las autoevaluaciones

3.1 **1.** ***a)*** $\bar{X} = \frac{\Sigma X}{n}$

b) $\bar{X} = \frac{\$267\ 100}{4} = \$66\ 775$

c) Estadístico, pues se trata de un valor muestral.

d) $66 775. La media de la muestra constituye nuestra mejor aproximación de la media poblacional.

2. ***a)*** $\mu = \frac{\Sigma X}{N}$

b) $\mu = \frac{498}{6} = 83$

c) Parámetro, porque se calculó empleando todos los valores de la población.

3.2 ***a)*** $237, calculado de la siguiente manera:

$$\frac{(95 \times \$400) + (126 \times \$200) + (79 \times \$100)}{95 + 126 + 79} = \$237.00$$

b) La ganancia por traje es de $12, que se determina mediante la operación $237 – costo de $200 – $25 de comisión. La ganancia total en los 300 trajes es de $3 600, la cual se calcula multiplicando 300 × $12.

3.3 **1.** ***a)*** $878

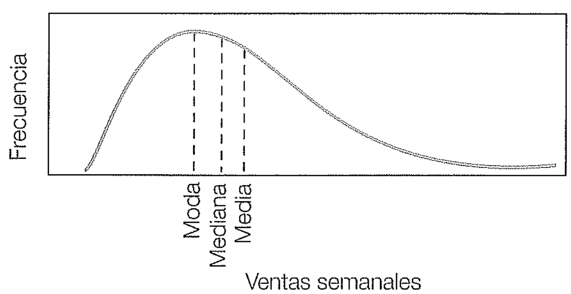
b) 3,3

2. ***a)*** 7, que se calcula mediante la operación (6 + 8)/2 = 7

b) 3,3

c) 0

3.4 ***a)***



b) Con sesgo positivo, ya que la media es el promedio más grande y la moda es el más pequeño.

3.5 **1.** ***a)*** Alrededor de 9.9%, que se obtiene con la raíz $\sqrt[4]{1.458602236}$.

b) Alrededor de 10.095%

c) Mayor que, por que 10.095 > 9.9.

2. 8.63%, que se determina mediante la operación

$$\sqrt[20]{\frac{120\ 520}{23\ 000}} - 1 = 1.0863 - 1$$

3.6 ***a)*** 22 000 de libras, que se determina restando 112 – 90

b) $\bar{X} = \frac{824}{8} = 103$ miles de libras

c)

X	$\lvert X - \bar{X} \rvert$	Desviación absoluta
95	$\lvert -8 \rvert$	8
103	$\lvert 0 \rvert$	0
105	$\lvert +2 \rvert$	2
110	$\lvert +7 \rvert$	7
104	$\lvert +1 \rvert$	1
105	$\lvert +2 \rvert$	2
112	$\lvert +9 \rvert$	9
90	$\lvert -13 \rvert$	13
		Total 42

$MD = \frac{42}{8} = 5.25$ miles de libras

3.7 ***a)*** $\mu = \frac{\$16\ 900}{5} = \$3\ 380$

b)
$$\sigma^2 = \frac{(3\ 536 - 3\ 380)^2 + \cdots + (3\ 622 - 3\ 380)^2}{5}$$
$$= \frac{(156)^2 + (-207)^2 + (68)^2 + (-259)^2 + (242)^2}{5}$$
$$= \frac{197\ 454}{5} = 39\ 490.8$$

c) $\sigma = \sqrt{39\ 490.8} = 198.72$

d) Hay más variación en la oficina de Pittsburgh, ya que la desviación estándar es mayor. La media también es mayor en la oficina de Pittsburgh.

3.8 2.33, que se calcula de la siguiente manera:

$$\bar{X} = \frac{\Sigma X}{n} = \frac{28}{7} = 4$$

X	$X - \bar{X}$	$(X - \bar{X})^2$
4	0	0
2	−2	4
5	1	1
4	0	0
5	1	1
2	−2	4
6	2	4
28	0	14

$$S^2 = \frac{\Sigma(X - \bar{X})^2}{n - 1} = \frac{14}{7 - 1} = 2.33$$

$$s = \sqrt{2.33} = 1.53$$

3.9 **a)** $k = \dfrac{14.15 - 14.00}{.10} = 1.5$

$k = \dfrac{13.85 - 14.0}{.10} = -1.5$

$1 - \dfrac{1}{(1.5)^2} = 1 - .44 = .56$

b) 13.8 y 14.2

3.10 **a)** Distribución de frecuencias.

b)

f	M	fM	$(M - \bar{X})$	$f(M - \bar{X})^2$
1	4	4	−8.2	67.24
4	8	32	−4.2	70.56
10	12	120	−0.2	0.40
3	16	48	3.8	43.32
2	20	40	7.8	121.68
20		244		303.20

$$\bar{X} = \frac{\Sigma fM}{M} = \frac{\$244}{20} = \$12.20$$

c) $s = \sqrt{\dfrac{303.20}{20 - 1}} = \$3.99$