11. En una prueba sobre consumo de gasolina se examinaron a 13 automóviles en un recorrido de 100 millas, tanto en ciudad como en carretera. Se obtuvieron los datos siguientes de rendimiento en millas por galón.

Ciudad:	16.2	16.7	15.9	14.4	13.2	15.3	16.8	16.0	16.1	15.3	15.2	15.3	16.2
Carretera:	19.4	20.6	18.3	18.6	19.2	17.4	17.2	18.6	19.0	21.1	19.4	18.5	18.7

Use la media, la mediana y la moda para indicar cuál es la diferencia en el consumo entre ciudad y carretera.

12. La empresa Walt Disney compró en 7.4 mil millones de dólares Pixar Animation Studios Inc. (*CNNMoney.com* 24 de enero de 2006). A continuación se presentan las películas animadas producidas por cada una de estas empresas (Disney y Pixar). Las ganancias están en millones de dólares. Calcule las ganancias totales, la media, la mediana y los cuartiles para comparar el éxito de las películas producidas por ambas empresas. ¿Sugieren dichos estadísticos por lo menos una razón por la que Disney haya podido estar interesada en comprar Pixar? Analice.

archivo en CD
Disney

Películas de Disney	Ganancias (millones de $)	Películas de Pixar	Ganancias (millones de $)
Pocahontas	346	*Toy Story*	362
Hunchback of Notre Dame	325	*A Bug's Life*	363
Hercules	253	*Toy Story 2*	485
Mulan	304	*Monsters, Inc.*	525
Tarzan	448	*Finding Nemo*	865
Dinosaur	354	*The Incredibles*	631
The Emperor's New Groove	169		
Lilo & Stitch	273		
Treasure Planet	110		
The Jungle Book 2	136		
Brother Bear	250		
Home on the Range	104		
Chicken Little	249		

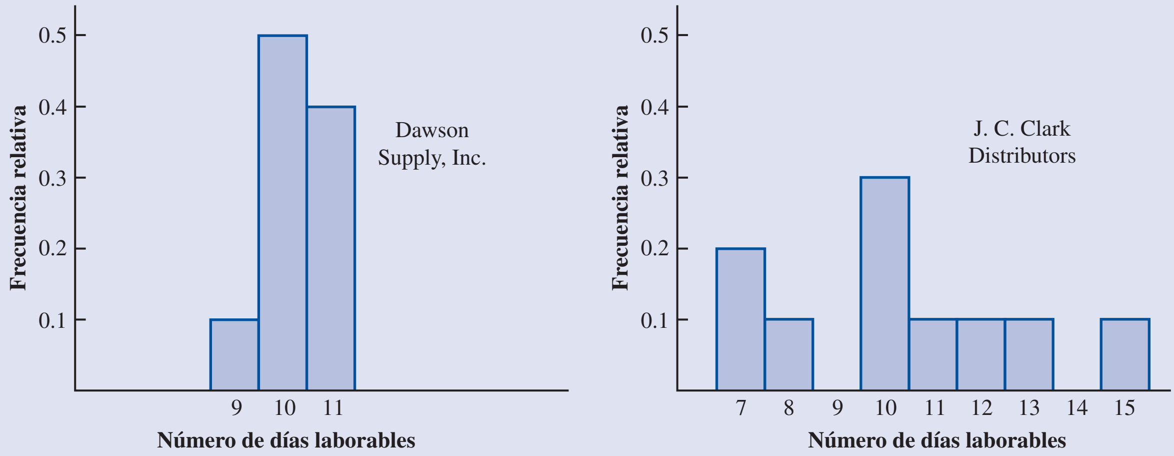
3.2 Medidas de variabilidad

La variabilidad en los tiempos de entrega produce incertidumbre en la planeación de la producción. Los métodos que se presentan en esta sección ayudan a medir y entender la variabilidad.

Además de las medidas de localización, suele ser útil considerar las medidas de variabilidad o de dispersión. Suponga que usted es el encargado de compras de una empresa grande y que con regularidad envía órdenes de compra a dos proveedores. Después de algunos meses de operación, se percata de que el número promedio de días que ambos proveedores requieren para surtir una orden es 10 días. En la figura 3.2 se presentan los histogramas que muestran el número de días que cada uno de los proveedores necesita para surtir una orden. Aunque en ambos casos este número promedio de días es 10 días, ¿muestran los dos proveedores el mismo grado de confiabilidad en términos de tiempos para surtir los productos? Observe la dispersión, o variabilidad, de estos tiempos en ambos histogramas. ¿Qué proveedor preferiría usted?

Para la mayoría de las empresas es importante recibir a tiempo los materiales que necesitan para sus procesos. En el caso de J. C. Clark Distributors sus tiempos de entrega, de siete u ocho días, parecen muy aceptables; sin embargo, sus pocos tiempos de entrega de 13 a 15 días resul-

FIGURA 3.2 DATOS HISTÓRICOS QUE MUESTRAN EL NÚMERO DE DÍAS REQUERIDOS PARA COMPLETAR UNA ORDER



tan desastrosos en términos de mantener ocupada a la fuerza de trabajo y de cumplir con el plan de producción. Este ejemplo ilustra una situación en que la variabilidad en los tiempos de entrega puede ser la consideración más importante en la elección de un proveedor. Para la mayor parte de los encargados de compras, la poca variabilidad que muestra en los tiempos de entrega de Dawson Supply, Inc. hará de esta empresa el proveedor preferido.

Ahora mostramos el estudio de algunas de las medidas de variabilidad más usadas.

Rango

La medida de variabilidad más sencilla es el **rango**.

RANGO

$$\text{Rango} = \text{Valor mayor} - \text{Valor menor}$$

De regreso a los datos de la tabla 3.1 sobre sueldos iniciales de los recién egresados de la carrera de administración, el mayor sueldo inicial es 3925 y el menor 3310. El rango es $3925 - 3310 = 615$.

Aunque el rango es la medida de variabilidad más fácil de calcular, rara vez se usa como única medida. La razón es que el rango se basa sólo en dos observaciones y, por tanto, los valores extremos tienen una gran influencia sobre él. Suponga que uno de los recién egresados haya tenido \$10 000 como sueldo inicial, entonces el rango será $10\,000 - 3310 = 6690$ en lugar de 615. Un valor así no sería muy descriptivo de la variabilidad de los datos ya que 11 de los 12 sueldos iniciales se encuentran entre 3310 y 3730.

Rango intercuartílico

Una medida que no es afectada por los valores extremos es el **rango intercuartílico (RIC)**. Esta medida de variabilidad es la diferencia entre el tercer cuartil Q_3 y el primer cuartil Q_1. En otras palabras, el rango intercuartílico es el rango en que se encuentra el 50% central de los datos.

RANGO INTERCUARTÍLICO

$$\text{IQR} = Q_3 - Q_1 \quad \textbf{(3.3)}$$

En los datos de los sueldos mensuales iniciales, los cuartiles son $Q_3 = 3600$ y $Q_1 = 3465$. Por lo tanto el rango intercuartílico es $3600 - 3465 = 135$.

Varianza

La **varianza** es una medida de variabilidad que utiliza todos los datos. La varianza está basada en la diferencia entre el valor de cada observación (x_i) y la media. A la diferencia entre cada valor x_i y la media ($\bar{x}$ cuando se trata de una muestra, μ cuando se trata de una población) se le llama *desviación respecto de la media.* Si se trata de una muestra, una desviación respecto de la media se escribe $(x_i - \bar{x})$, y si se trata de una población se escribe $(x_i - \mu)$. Para calcular la varianza, estas desviaciones respecto de la media *se elevan al cuadrado*.

Si los datos son de una población, el promedio de estas desviaciones elevadas al cuadrado es la *varianza poblacional.* La varianza poblacional se denota con la letra griega σ^2. En una población en la que hay N observaciones y la media poblacional es μ, la varianza poblacional se define como sigue.

VARIANZA POBLACIONAL

$$\sigma^2 = \frac{\Sigma(x_i - \mu)^2}{N} \quad \textbf{(3.4)}$$

En la mayor parte de las aplicaciones de la estadística, los datos a analizar provienen de una muestra. Cuando se calcula la varianza muestral, lo que interesa es estimar la varianza poblacional σ^2. Aunque una explicación detallada está más allá del alcance de este libro, es posible demostrar que si la suma de los cuadrados de las desviaciones respecto de la media se divide entre $n - 1$, en lugar de entre n, la varianza muestral que se obtiene constituye un estimador no sesgado de la varianza poblacional. Por esta razón, la *varianza muestral*, que se denota por s^2, se define como sigue.

La varianza muestral s^2 es el estimador de la varianza poblacional σ^2 .

VARIANZA MUESTRAL

$$s^2 = \frac{\Sigma(x_i - \bar{x})^2}{n - 1} \quad \textbf{(3.5)}$$

Para ilustrar el cálculo de la varianza muestral, se emplean los datos de los tamaños de cinco grupos de una universidad, presentados en la sección 3.1. En la tabla 3.3 aparece un resumen de los datos con el cálculo de las desviaciones respecto de la media y de los cuadrados de las desviaciones respecto de la media. La suma de los cuadrados de las desviaciones respecto de la media es $\Sigma(x_i - \bar{x})^2 = 256$. Por tanto, siendo $n - 1 = 4$, la varianza muestral es

$$s^2 = \frac{\Sigma(x_i - \bar{x})^2}{n - 1} = \frac{256}{4} = 64$$

Antes de continuar, hay que hacer notar que las unidades correspondientes a la varianza muestral suelen causar confusión. Como los valores que se suman para calcular la varianza, $(x_i - \bar{x})^2$, están elevados al cuadrado, las unidades correspondientes a la varianza muestral tam-

TABLA 3.3 CÁLCULO DE LAS DESVIACIONES Y DE LOS CUADRADOS DE LAS DESVIACIONES RESPECTO DE LA MEDIA EMPLEANDO LOS DATOS DE LOS TAMAÑOS DE CINCO GRUPOS DE ESTADOUNIDENSES

Número de estudiantes en un grupo (x_i)	Número promedio de alumnos en un grupo ($\bar{x}$)	Desviación respecto a la media ($x_i - \bar{x}$)	Cuadrado de la desviación respecto de la media $(x_i - \bar{x})^2$
46	44	2	4
54	44	10	100
42	44	−2	4
46	44	2	4
32	44	−12	144
		0	256
		$\Sigma(x_i - \bar{x})$	$\Sigma(x_i - \bar{x})^2$

bién están *elevadas al cuadrado*. Por ejemplo, la varianza muestral en los datos de la cantidad de alumnos en los grupos es $s^2 = 64$ (estudiantes)^2. Las unidades al cuadrado de la varianza dificultan la comprensión e interpretación intuitiva de los valores numéricos de la varianza. Aquí lo recomendable es entender la varianza como una medida útil para comparar la variabilidad de dos o más variables. Al comparar variables, la que tiene la varianza mayor, muestra más variabilidad. Otra interpretación del valor de la varianza suele ser innecesaria.

La varianza sirve para comparar la variabilidad de dos o más variables.

Para tener otra ilustración del cálculo de la varianza muestral, considere los sueldos iniciales de 12 recién egresados de la carrera de administración, presentados en la tabla 3.1. En la sección 3.1 se vio que la media muestral de los sueldos mensuales iniciales era 3540. En la tabla 3.4 se muestra el cálculo de la varianza muestral ($s^2 = 27\,440.91$).

TABLA 3.4 CÁLCULO DE LA VARIANZA MUESTRAL CON LOS DATOS DE LOS SUELDOS INICIALES

Sueldo mensual (x_i)	Media muestral ($\bar{x}$)	Desviación respecto de la media ($x_i - \bar{x}$)	Cuadrado de la desviación respecto de la media $(x_i - \bar{x})^2$
3450	3540	−90	8 100
3550	3540	10	100
3650	3540	110	12 100
3480	3540	−60	3 600
3355	3540	−185	34 225
3310	3540	−230	52 900
3490	3540	−50	2 500
3730	3540	190	36 100
3540	3540	0	0
3925	3540	385	148 225
3520	3540	−20	400
3480	3540	−60	3 600
		0	301 850
		$\Sigma(x_i - \bar{x})$	$\Sigma(x_i - \bar{x})^2$

Empleando la ecuación (3.5),

$$s^2 = \frac{\Sigma(x_i - \bar{x})^2}{n - 1} = \frac{301\,850}{11} = 27\,440.91$$

En las tablas 3.3 y 3.4 se presenta la suma, tanto de las desviaciones respecto de la media como de los cuadrados de las desviaciones respecto de la media. En todo conjunto de datos, la suma de las desviaciones respecto de la media será *siempre igual a cero.* Observe que en las tablas 3.3 y 3.4 $\Sigma(x_i - \bar{x}) = 0$. Las desviaciones positivas y las desviaciones negativas se anulan mutuamente haciendo que la suma de las desviaciones respecto a la media sea igual a cero.

Desviación estándar

La **desviación estándar** se define como la raíz cuadrada positiva de la varianza. Continuando con la notación adoptada para la varianza muestral y para la varianza poblacional, se emplea s para denotar la desviación estándar muestral y σ para denotar la desviación estándar poblacional. La desviación estándar se obtiene de la varianza como sigue.

La desviación estándar muestral s es el estimador de la desviación estándar poblacional σ.

DESVIACIÓN ESTÁNDAR

$$\text{Desviación estándar muestral} = s = \sqrt{s^2} \tag{3.6}$$

$$\text{Desviación estándar poblacional} = \sigma = \sqrt{\sigma^2} \tag{3.7}$$

Recuerde que la varianza muestral para los tamaños de cinco grupos de una universidad es $s^2 = 64$. Por tanto, la desviación estándar muestral es $s = \sqrt{64} = 8$. En los datos de los sueldos iniciales, la desviación estándar es $s = \sqrt{27\,440.91} = 165.65$.

La desviación estándar es más fácil de interpretar que la varianza debido a que la desviación estándar se mide en las mismas unidades que los datos.

¿Qué se gana con convertir la varianza en la correspondiente desviación estándar? Recuerde que en la varianza las unidades están elevadas al cuadrado. Por ejemplo, la varianza muestral de los datos de los sueldos iniciales de los egresados de administración es $s^2 = 27{,}440.91$ (dólares)^2. Como la desviación estándar es la raíz cuadrada de la varianza, las unidades de la varianza, dólares al cuadrado, se convierten en dólares en la desviación estándar. Por tanto, la desviación estándar de los sueldos iniciales es $165.65. En otras palabras, la desviación estándar se mide en las mismas unidades que los datos originales. Por esta razón es más fácil comparar la desviación estándar con la media y con otros estadísticos que se miden en las mismas unidades que los datos originales.

Coeficiente de variación

El coeficiente de variación es una medida relativa de la variabilidad; mide la desviación estándar en relación con la media.

En algunas ocasiones se requiere un estadístico descriptivo que indique cuán grande es la desviación estándar en relación con la media. Esta medida es el **coeficiente de variación** y se representa como porcentaje.

COEFICIENTE DE VARIACIÓN

$$\left(\frac{\text{Desviación estándar}}{\text{Media}} \times 100\right)\% \tag{3.8}$$

En los datos de los tamaños de los cinco grupos de estudiantes, se encontró una media muestral de 44 y una desviación estándar muestral de 8. El coeficiente de variación es $[(8/44) \times 100]\% = 18.2\%$. Expresado en palabras, el coeficiente de variación indica que la desviación estándar muestral es 18.2% del valor de la media muestral. En los datos de los sueldos iniciales, la media muestral encontrada es 3540 y la desviación estándar muestral es 165.65, el coeficiente de variación, $[(165.65/3540) \times 100]\% = 4.7\%$, indica que la desviación estándar muestral es sólo 4.7% del valor de la media muestral. En general, el coeficiente de variación es un estadístico útil para comparar la variabilidad de variables que tienen desviaciones estándar distintas y medias distintas.

NOTAS Y COMENTARIOS

1. Los paquetes de software para estadística y las hojas de cálculo sirven para buscar los estadísticos descriptivos presentados en este capítulo. Una vez que los datos se han ingresado en una hoja de cálculo, basta emplear unos cuantos comandos sencillos para obtener los estadísticos deseados. En los apéndices 3.1 y 3.2 se muestra cómo usar Minitab y Excel para lograrlo.
2. La desviación estándar suele usarse como medida del riesgo relacionado con una inversión en acciones o en fondos de acciones (*BussinesWeek,* 7 de enero de 2000). Proporciona una medida de cómo fluctúa la rentabilidad mensual respecto de la rentabilidad promedio a largo plazo.
3. Redondear los valores de la media muestral $\bar{x}$ y de los cuadrados de las desviaciones $(x_i - \bar{x})^2$ puede introducir errores cuando se emplea una calculadora para el cálculo de la varianza y de la desviación estándar. Para reducir los errores de redondeo se recomienda conservar por lo menos seis dígitos significativos en los cálculos intermedios. La varianza o la desviación estándar obtenidos se redondean entonces a menos dígitos significativos.
4. Otra fórmula alterna para el cálculo de la varianza muestral es

$$s^2 = \frac{\Sigma x_i^2 - n\bar{x}^2}{n - 1}$$

donde $\Sigma x_i^2 = x_1^2 + x_2^2 + \cdots + x_n^2$.

Ejercicios

Métodos

13. Considere una muestra con los datos 10, 20, 12, 17 y 16. Calcule el rango y el rango intercuartílico.
14. Considere una muestra que tiene como valores 10, 20, 12, 17 y 16. Calcule la varianza y la desviación estándar.

Auto examen

15. Considere una muestra con valores 27, 25, 0, 15, 30, 34, 28 y 25. Calcule el rango, el rango intercuartílico, la varianza y la desviación estándar.

Aplicaciones

Auto examen

16. Las puntuaciones obtenidas por un jugador de boliche en seis juegos fueron 182, 168, 184, 190, 170 y 174. Use estos datos como una muestra y calcule los estadísticos descriptivos siguientes
    a. Rango
    b. Varianza
    c. Desviación estándar
    d. Coeficiente de variación
17. A *home theater in a box* es la manera más sencilla y económica de tener sonido envolvente en un centro de entretenimiento en casa. A continuación se presenta una muestra de precios (*Consumer Report Buying Guide* 2004). Los precios corresponden a modelos con y sin reproductor de DVD.

Modelos con reproductor de DVD	Precio	Modelos sin reproductor de DVD	Precio
Sony HT-1800DP	$450	Pioneer HTP-230	$300
Pioneer HTD-330DV	300	Sony HT-DDW750	300
Sony HT-C800DP	400	Kenwood HTB-306	360
Panasonic SC-HT900	500	RCA RT-2600	290
Panasonic SC-MTI	400	Kenwood HTB-206	300

a. Calcule el precio medio de los modelos con reproductor de DVD y el precio medio de los modelos sin reproductor de DVD. ¿Cuánto es lo que se paga de más por tener un reproductor de DVD en casa?
b. Calcule el rango, la varianza y la desviación estándar de las dos muestras. ¿Qué le dice esta información acerca de los precios de los modelos con y sin reproductor de DVD?

18. Las tarifas de renta de automóviles por día en siete ciudades del este de Estados Unidos son las siguientes (*The Wall Street Journal* 16 de enero de 2004).

Ciudad	Tarifa por día
Boston	$43
Atlanta	35
Miami	34
New York	58
Orlando	30
Pittsburgh	30
Washington, D.C.	36

a. Calcule la media, la varianza y la desviación estándar de estas tarifas.
b. En una muestra similar de siete ciudades del oeste la media muestral de las tarifas fue de $38 por día. La varianza y la desviación estándar fueron 12.3 y 3.5 cada una. Analice la diferencia entre las tarifas de las ciudades del este y del oeste.

19. *Los Angeles Times* informa con regularidad sobre el índice de la calidad del aire en varias regiones del sur de California. En una muestra de los índices de calidad del aire en Pomona se tienen los datos siguientes: 28, 42, 58, 48, 45, 55, 60, 49 y 50.
a. Calcule el rango y el rango intercuartílico.
b. Calcule la varianza muestral y la desviación estándar muestral.
c. En una muestra de índices de calidad del aire en Anaheim, la media muestral es 48.5, la varianza muestral es 136 y la desviación estándar muestral es 11.66. Con base en estos estadísticos descriptivos compare la calidad del aire en Pomona y en Anaheim.

20. A continuación se presentan los datos que se usaron para elaborar los histogramas sobre el número de días necesarios para surtir una orden (véase la figura 3.2).

Días de entrega de Dawson Supply, Inc.:	11	10	9	10	11	11	10	11	10	10
Días de entrega de Clark Distributors:	8	10	13	7	10	11	10	7	15	12

Use el rango y la desviación estándar para sustentar la observación hecha antes de que Dawson Supply proporcione los tiempos de entrega más consistentes.

21. ¿Cómo están los costos de abarrotes en el país? A partir de una canasta alimenticia de 10 artículos entre los que se encuentran carne, leche, pan, huevos, café, papas, cereal y jugo de naranja, la revista *Where to Retire* calculó el costo de la canasta alimenticia en seis ciudades y en seis zonas con personas jubiladas en todo el país (*Where to Retire* noviembre/diciembre de 2003). Los datos encontrados, al dólar más cercano, se presentan a continuación.

Ciudad	Costo	Zona de jubilados	Costo
Buffalo, NY	$33	Biloxi-Gulfport, MS	$29
Des Moines, IA	27	Asheville, NC	32
Hartford, CT	32	Flagstaff, AZ	32
Los Angeles, CA	38	Hilton Head, SC	34
Miami, FL	36	Fort Myers, FL	34
Pittsburgh, PA	32	Santa Fe, NM	31

a. Calcule la media, varianza y desviación estándar de las ciudades y de las zonas de jubilados.
b. ¿Qué observaciones puede hacer con base en estas dos muestras?

archivo en CD
Broker

22. La Asociación Estadounidense de Inversionistas Individuales realiza cada año una investigación sobre los corredores de bolsa con descuento (*AAII Journal*, enero de 2003). En la tabla 3.2 se muestran las comisiones que cobran 24 corredores de bolsa con descuento por dos tipos de transacciones: transacción con ayuda del corredor de 100 acciones a $50 la acción y transacción en línea de 500 acciones a $50 la acción.
   a. Calcule el rango y el rango intercuartílico en cada tipo de transacción.
   b. Calcule la varianza y la desviación estándar en cada tipo de transacción.
   c. Calcule el coeficiente de variación en cada tipo de transacción.
   d. Compare la variabilidad en el costo que hay en los dos tipos de transacciones

24. Las puntuaciones de un jugador de golf en el 2005 y 2006 son las siguientes:

2005	74	78	79	77	75	73	75	77
2006	71	70	75	77	85	80	71	79

   a. Use la media y la desviación estándar para evaluar a este jugador de golf en estos dos años.
   b. ¿Cuál es la principal diferencia en su desempeño en estos dos años? ¿Se puede ver algún progreso en sus puntuaciones del 2006?, ¿cuál?

24. Los siguientes son los tiempos que hicieron los velocistas de los equipos de pista y campo de una universidad en un cuarto de milla y en una milla (los tiempos están en minutos).

Tiempos en un cuarto de milla:	0.92	0.98	1.04	0.90	0.99
Tiempos en una milla:	4.52	4.35	4.60	4.70	4.50

Después de ver estos datos, el entrenador comentó que en un cuarto de milla los tiempos eran más homogéneos. Use la desviación estándar y el coeficiente de variación para resumir la variabilidad en los datos. El uso del coeficiente de variación, ¿indica que la aseveración del entrenador es correcta?

3.3 Medidas de la forma de la distribución, de la posición relativa y de la detección de observaciones atípicas

Se han descrito ya varias medidas de localización y de variabilidad de los datos. Además de estas medidas se necesita una medida de la forma de la distribución. En el capítulo 2 se vio que un histograma es una representación gráfica que muestra la forma de una distribución. Una medida numérica importante de la forma de una distribución es el **sesgo**.

Forma de la distribución

En la figura 3.3 se muestran cuatro histogramas elaborados a partir de distribuciones de frecuencias relativas. Los histogramas A y B son moderadamente sesgados. El histograma A es sesgado a la izquierda, su sesgo es −0.85. El histograma B es sesgado a la derecha, su sesgo es +0.85. El histograma C es simétrico; su sesgo es cero. El histograma D es muy sesgado a la derecha; su sesgo es 1.62. La fórmula que se usa para calcular el sesgo es un poco complicada.* Sin embargo, es fácil de calcular empleando el software para estadística (véase los apéndices 3.1 y 3.2). En

*La fórmula para calcular el sesgo de datos muestrales es:

$$\text{Sesgo} = \frac{n}{(n-1)(n-2)} \sum \left(\frac{x_i - \bar{x}}{s} \right)^3$$