

3D-6. Supongamos que las siguientes son distribuciones de frecuencia de las edades de adultos mayores en dos centros de retiro diferentes.

- a) Construye polígonos de frecuencia sobrepuestos de estos datos. (Atención: los tamaños muestrales difieren.)
- b) ¿Qué revela el gráfico?

Edad	Centro I	Centro II
	f	f
67	6	2
68	5	1
69	5	0
70	3	1
71	4	2
72	4	3
73	3	2
74	2	2
75	1	4
76	1	3
77	2	5
78	1	3
79	0	4
80	2	3

APLICACIONES OPCIONALES DE COMPUTADORA PARA EL CAPÍTULO 3

En el sitio web www.mhhe.com/ritchev2, en *The Statistical Imagination*, están disponibles ejercicios computarizados del capítulo. Estos ejercicios indican cómo elaborar gráficos y tablas con el uso de *SPSS for Windows* y cómo seleccionar estilos gráficos apropiados. Además, el apéndice D de este texto, *An Introduction to SPSS*, proporciona instrucciones básicas para elaborar gráficos y tablas.

Como es el caso para software gráfico, los gráficos en *SPSS* tienen características preestablecidas (ajustes por *default*) que pueden construir un gráfico diferente del que se pretenda. Por tanto, las salidas impresas gráficas suelen requerir editarse. Es oportuno un aviso de precaución. Trata a un programa computarizado de gráficos simplemente como herramienta de dibujo. Procura examinar con todo cuidado un gráfico para cerciorarte de que es preciso. Sigue los Lineamientos de Construcción de Gráficos y Tablas que se ven al inicio de este capítulo. El lector, y no el paquete de software, es responsable en última instancia del producto final.

Estimación de promedios

RESUMEN DEL CAPÍTULO

Introducción	107	Curvas de distribución de frecuencias: relaciones entre la media, la mediana y la moda	118
La media	108	La distribución normal	118
Pensamiento proporcional sobre la media	109	Distribuciones sesgadas	119
Debilidades potenciales de la media: situaciones en las que reportarla sola puede conducir a errores	111	Uso de los datos de una muestra para estimar la forma de una distribución de puntuaciones en una población	120
La mediana	112	Organización de los datos para calcular los estadísticos de tendencia central	122
Debilidades potenciales de la mediana: situaciones en las que reportarla sola puede conducir a errores	114	Formato de hoja de cálculo para calcular estadísticos de tendencia central	122
La moda	115	Formato de distribución de frecuencias para calcular la moda	123
Debilidades potenciales de la moda: situaciones en las que reportarla sola puede conducir a errores	116	Insensatez y falacias estadísticas: mezcla de subgrupos en el cálculo de la media	124
Estadísticos de tendencia central y el nivel apropiado de medición	117		

Introducción

Todos estamos familiarizados con el concepto general de promedio, en situaciones tales como una calificación promedio, un ingreso promedio, una puntuación promedio en el boliche o un promedio de bateo. Si alguien tiene un "promedio" de alguna manera, por ejemplo altura, peso, inteligencia, etc., esta persona no es atípica. Poseer un promedio significa ser como la mayoría de las personas.

En una distribución de puntuaciones, un promedio caerá entre las puntuaciones extremas, en alguna parte del área media de la distribución de puntuaciones. Por ejemplo, la mayoría de los hombres no son demasiado altos o bajos, están "sobre el promedio". A esta puntuación típica o promedio la llamamos la **tendencia central** de la variable. Un **estadístico de tendencia central** proporciona una estimación de la puntuación típica, común o normal encontrada en una distribución de puntuaciones en bruto. Por ejemplo, las estaturas de los hombres estadounidenses tienden a agruparse alrededor de cinco pies con ocho pulgadas, y

los pequeños saludables pesan alrededor de siete libras al nacer. Si Bob tiene un promedio de 165 en el boliche, no esperamos que obtenga esta puntuación exacta en cada juego, pero conseguirá cercanamente esa puntuación la mayoría de las veces.

Estadístico de tendencia central Estadístico que proporciona una estimación de la puntuación típica, común o normal encontrada en una distribución de puntuaciones en bruto.

Existen tres estadísticos de tendencia central comunes: la media, la mediana y la moda. ¿Por qué tres? Porque cada uno tiene aspectos fuertes, pero también debilidades potenciales, dependiendo de la forma particular de la distribución de puntuaciones de una variable. Según sea la forma de una distribución, una medición del promedio puede resultar más exacta que otra, y, en ocasiones, informar cualquier estadístico de tendencia central sólo conduciría a errores o no proporcionaría información suficiente.

La media

La media aritmética de una distribución de puntuaciones (o, simplemente la media) consiste en un estadístico de tendencia central que es familiar a cualquier estudiante que haya calculado el promedio de sus calificaciones para algún curso. La **media** es la suma de todas las puntuaciones dividida entre el número de puntuaciones observadas (es decir, el tamaño de la muestra). Para calcular la media de una variable, simplemente sumamos todas las puntuaciones y dividimos el resultado entre el tamaño de la muestra.

La media Suma de todas las puntuaciones dividida entre el número de puntuaciones observadas (es decir, el tamaño de la muestra).

La media es el estadístico de tendencia central más útil. Con un cálculo matemático rápido, ofrece un resumen de las puntuaciones típicas o promedio en una distribución. Puesto que emplea la operación matemática de división, la media se aplica a las variables de intervalo/razón.

En las fórmulas matemáticas el símbolo convencional utilizado para representar el nombre de una variable es una letra mayúscula. Las letras X y Y se emplean con frecuencia. Por ejemplo, podríamos emplear X para simbolizar la edad y Y para la altura. A menudo, Y se usa para la variable dependiente y X para la variable independiente. Por ejemplo, pondríamos Y = calificación promedio (CP) de la universidad con el siguiente conjunto de variables predictoras: X_1 = calificación promedio (CP) de la preparatoria, X_2 = puntuaciones en el examen de admisión a la universidad, X_3 = habilidad en la comprensión de lectura y X_4 = año de escolaridad.

Para una variable X , cualquier cosa que definamos, el símbolo para la media calculada con datos de la muestra es $\bar{X}$, que se llama "X barra". Por ejemplo, si X = edad, y la edad media del grupo de estadística es 20.5 años, decimos "X barra es igual a 20.5 años". Recuerda especificar las unidades de medida de la variable, en este caso, años. La media se calcula como sigue (Σ se lee como "la suma").

Cálculo de la media

$$\bar{X} = \frac{\Sigma X}{n}$$

donde

$\bar{X}$ = la media de la variable X de intervalo/razón con datos de la muestra

ΣX = la suma de todas las puntuaciones individuales para la variable X

n = el número de observaciones (es decir, el tamaño muestral)

Si hay 12 niños en una muestra, cuyas edades son 6, 12, 5, 10, 9, 10, 8, 7, 9, 11, 8 y 10 años, su edad media es

$$\begin{aligned}\bar{X} &= \frac{\Sigma X}{n} = \frac{6 + 12 + 5 + 10 + 9 + 10 + 8 + 7 + 9 + 11 + 8 + 10}{12} \\ &= \frac{105 \text{ años}}{12} = 8.75 \text{ años}\end{aligned}$$

Técnicamente, la media es 8.75 años *por niño*, pero omitimos la unidad del denominador. Conceptualmente, el valor de la media nos dice cuáles serían las puntuaciones X de una muestra si es que cada sujeto de la muestra tuviera la misma puntuación. En el ejemplo anterior, 8.75 años (es decir, 8 años nueve meses) sería la edad de cada niño si todos los niños tuvieran exactamente la misma edad. Es útil, entonces, pensar en la media como una medición de "partes iguales". Por ejemplo, si quisiéramos saber la cantidad media de dinero en efectivo que llevan consigo los estudiantes de un salón de clases, pondríamos todo el dinero en efectivo en un recipiente y lo dividiríamos equitativamente. (¿Algún voluntario?) La cantidad que recibiría cada persona sería el valor medio del dinero en efectivo. La media también puede ser considerada como un punto de equilibrio, es decir, el punto en el que se equilibran las diferencias entre la media de X y las puntuaciones individuales X de la distribución. En el capítulo 5 ampliaremos esta noción.

Por último, al calcular los estadísticos de tendencia central, particularmente la media, debe tenerse cuidado para no incluir las puntuaciones codificadas como casos perdidos. Al determinar la media sólo se incluyen los casos "válidos". Por ejemplo, si en una muestra de 49 personas 2 de ellas no informaron sus edades, la suma de las edades se dividiría entre 47, que es el número de puntuaciones válidas, en lugar de dividirla entre el tamaño de la muestra 49. Además, con archivos de computadora, debe tenerse cuidado de no sumar los códigos de "valor perdido" (como 999) a la suma de las puntuaciones.

Pensamiento proporcional sobre la media

Combinación de las medias de dos muestras de tamaño diferente La media es el estadístico de tendencia central más ampliamente usado de variables de intervalo/razón. Así, es importante que tengamos un buen sentido de proporción respecto de su cálculo. Primero,

examinemos una situación donde se comete un error común: combinar las medias de dos grupos sumando las dos medias y dividiendo el resultado entre 2. [El único momento en que esto no representa un error es cuando los dos grupos tienen los mismos tamaños de muestra (es decir, cuando las n son iguales).] Por ejemplo, observa el número medio de días de vacaciones por año (X) para el grupo 1, las ocho secretarías de un banco local, y, para el grupo 2, los tres vicepresidentes. Para las ocho secretarías:

$$\begin{aligned}\bar{X}_{(\text{grupo1})} &= \frac{\Sigma X_{(\text{grupo1})}}{n_{(\text{grupo1})}} = \frac{7 + 10 + 7 + 12 + 16 + 7 + 14 + 10}{8} \\ &= \frac{83 \text{ días}}{8} = 10.38 \text{ días de vacaciones}\end{aligned}$$

Para los tres vicepresidentes:

$$\begin{aligned}\bar{X}_{(\text{grupo2})} &= \frac{\Sigma X_{(\text{grupo2})}}{n_{(\text{grupo2})}} = \frac{60 + 30 + 30}{3} \\ &= \frac{120 \text{ días}}{3} = 40.00 \text{ días de vacaciones}\end{aligned}$$

Si calculamos *incorrectamente* la media de la oficina completa sumando estas dos medias y dividiendo el resultado entre 2, obtendríamos la respuesta errónea de 25.19 días de vacaciones. El cálculo correcto para esta media combinada es

$$\begin{aligned}\bar{X}_{(\text{grupos 1 y 2 combinados})} &= \frac{\Sigma X_{(\text{grupo1})} + \Sigma X_{(\text{grupo2})}}{n_{(\text{grupo1})} + n_{(\text{grupo2})}} \\ &= \frac{83 + 120}{8 + 3} = \frac{203}{11} = 18.45 \text{ días de vacaciones}\end{aligned}$$

Analizando un poco veremos que esta formulación es equivalente a tratar a los 11 empleados como una muestra. Para ejemplificar casos al "promediar" erróneamente las medias de un grupo, véase el apartado de "Insensatez y falacias estadísticas" al final de este capítulo.

Cálculo de la media combinada de dos muestras de tamaño diferente

Dadas las medias y tamaños muestrales de dos grupos:

$$\bar{X}_{(\text{grupos 1 y 2 combinados})} = \frac{\Sigma X_{(\text{grupo1})} + \Sigma X_{(\text{grupo2})}}{n_{(\text{grupo1})} + n_{(\text{grupo2})}}$$

donde

$\bar{X}$ = la media de la variable de intervalo/razón, X , calculada en datos muestrales

$$\Sigma X_{(\text{grupo})} = (n)_{(\text{grupo})}(\bar{X})_{(\text{grupo})}$$

y

n = el número de observaciones (es decir, el tamaño muestral)

Ejemplo: supongamos X = calificación en examen final; la calificación media para los 13 estudiantes del último año del grupo es 87, y la calificación media para los 16 estudiantes del penúltimo año es 79. ¿Cuál es la calificación media para los dos grupos combinados?

1. Calcula la ΣX para cada grupo:

$$\Sigma X_{(\text{último año})} = (13)(87) = 1\,131 \text{ puntos de examen}$$

$$\Sigma X_{(\text{penúltimo})} = (16)(79) = 1\,264 \text{ puntos de examen}$$

2. Sustituye las sumas en la ecuación precedente:

$$\begin{aligned}\bar{X}_{(\text{de último y penúltimo})} &= \frac{\Sigma X_{(\text{último})} + \Sigma X_{(\text{penúltimo})}}{n_{(\text{último})} + n_{(\text{penúltimo})}} \\ &= \frac{1\,131 + 1\,264}{13 + 16} = \frac{2\,395}{29} = 82.59 \text{ puntos de examen}\end{aligned}$$

Debilidades potenciales de la media: situaciones en las que reportarla sola puede conducir a errores

Cuando se reporta un estadístico de tendencia central, tendemos a suponer que su valor es representativo de puntuaciones típicas en la parte central de una distribución. En ocasiones, sin embargo, cuando se informa la media puede conducir a errores al respecto. Éste es el caso porque el cálculo de la media puede inflarse (aumentarse) o desinflarse (disminuir) debido a puntuaciones o valores extremos. Puntuaciones muy altas, o valores extremos positivos, inflan el valor de la media "agrandando" la suma de X (es decir ΣX) en el numerador de la fórmula. Puntuaciones sumamente bajas en una distribución, o valores extremos negativos, desinflan el valor de la media "encogiéndolo" ΣX . Por ejemplo, suponga que calculamos la cantidad media del dinero en efectivo que llevan 10 estudiantes. Idealmente, esta media debe indicarnos cuál es la cantidad típica. Pero supongamos que un estudiante cobró un cheque por \$400 y nuestro cálculo es el siguiente, donde X = la cantidad de dinero en efectivo de cada estudiante (para simplificar, se redondea al dólar más cercano):

$$\begin{aligned}\bar{X} &= \frac{\Sigma X}{n} = \frac{5 + 2 + 6 + 10 + 8 + 3 + 9 + 11 + 5 + 400}{10} \\ &= \frac{\$459}{10} = \$45.90 \approx \$46\end{aligned}$$

Por obvias razones, esta media de \$46 no representa la cantidad de dinero promedio típica, o la tendencia central que los alumnos pueden llevar en efectivo. La mayoría de los estudiantes tiene menos de \$10, y reportar una media de \$46 es engañoso. El cálculo de la media se distorsiona por la presencia de un valor aislado. Para obtener un sentido de proporción sobre cómo se calcula la fórmula de la media, examina la relación entre el numerador (ΣX) y el denominador (n). Cuando ΣX es grande y n es pequeño, la media será grande. Cuando ΣX es grande debido a la presencia de uno o dos valores extremos de alto valor, la media se "inflará" hasta un valor grande.

Recuerda que nuestro objetivo es usar estadísticos de muestras para estimar los parámetros de una población. Si se reporta una media *muestral* inflada o disminuida, se presentará un resumen distorsionado de las puntuaciones que obtienen los sujetos de una población. Esta limitación de la media es un problema especial con muestras pequeñas; cuanto menor sea la muestra, mayor será la distorsión que genere un valor extremo. Por ejemplo, calcula la edad media de la siguiente muestra de cinco estudiantes de la universidad local, donde un estudiante de la muestra tiene una edad extremadamente alta: 19, 19, 20, 21, y 54 años. La respuesta dejará la impresión que esta muestra está bastante arriba de la edad típica en la universidad, cuando, en realidad, cuatro de los cinco estudiantes *tienen* la edad típica. También observa lo que sucede cuando existe una puntuación sumamente baja, como con esta muestra de edades: 8, 19, 20 y 21 años. En tales casos, los valores extremos deben eliminarse y la media debe calcularse de nuevo sin ellos. Al reportar esta "media ajustada", notamos por qué se realizó el ajuste.

En cualquier momento que calculemos una media, en especial con una muestra pequeña, primero examinamos la distribución de frecuencias de la puntuación en bruto para los valores extremos. Un recurso práctico para esto es un gráfico de caja (capítulo 3). Ya que la media es más útil que la mediana y la moda, con frecuencia ajustamos las puntuaciones de una distribución para reducir los efectos de los valores extremos en su cálculo. Los efectos deformantes de los valores extremos se mencionan en todo el texto.

La mediana

La **mediana** (Mdn) es la puntuación central en una distribución ordenada, es decir, el valor de una variable que divide en mitades a la distribución de las puntuaciones, la puntuación por arriba de la cual queda la mitad de los casos y por debajo queda la otra mitad. Por ejemplo, si la media del ingreso familiar en la ciudad Cornbelt es \$26 000, la mitad de las familias de esta ciudad tienen ingresos mayores a \$26 000 y la otra mitad ingresos menores a \$26 000. Conceptualmente, la mediana es un punto de localización, la puntuación de la mitad. La mediana trae a colación una posición geográfica entre áreas iguales, como la mediana de una carretera. La puntuación mediana también es igual al percentil 50, el punto bajo en el cual caen el 50% de las observaciones. Entre los tres estadísticos de tendencia central, la mediana es más útil cuando una distribución está sesgada (es decir, tiene pocas puntuaciones hacia un lado). Por ejemplo, la mediana del precio de las ventas recientes de viviendas es preferible a la media del precio, porque unas cuantas ventas de alto precio incrementarían el valor de la media.

La mediana Para una variable ordinal o de intervalo/razón, es la puntuación central en una distribución ordenada, la puntuación que deja debajo de sí la mitad de los casos y, por arriba, la otra mitad.

Para calcular la mediana de una distribución, primero debes ordenar las puntuaciones para una variable X , es decir, las puntuaciones **deben** colocarse en orden de tamaño, de menor a mayor o de mayor a menor. Divide entre 2 el tamaño de la muestra n para acercarte a la puntuación de la mitad de la distribución. Si n es un número impar, la mediana será un caso real en la muestra. Supongamos, por ejemplo, que tenemos una muestra de cinco familias con los siguientes ingresos mensuales (X):

Ubicación de la Mdn					
			↓		
Orden de los casos	1	2	3	4	5
Valores de X	\$3 540	\$4 675	\$7 350	\$9 860	\$19 000
			↑		
Mdn = un valor de X					

La mediana de ingreso es \$7 350, el valor de X para la tercera puntuación ordenada.

Si n es un número par, la mediana se localiza entre las dos puntuaciones de la mitad y se calcula tomando la media de esas dos puntuaciones. Por ejemplo, si una sexta familia con ingreso de \$20 000 se inserta en la muestra anterior,

Ubicación de la Mdn						
			↓			
Orden de los casos	1	2	3	4	5	6
Valores de X	\$3 540	\$4 675	\$7 350	\$9 860	\$19 000	\$20 000
				↑		
Mdn = \$8 605						

La mediana se sitúa entre el tercero y cuarto casos. Se calcula sumando las puntuaciones de \$7 350 y \$9 860 y dividiendo entre 2.

Con una muestra pequeña, localizar la mediana es un trabajo sencillo; con una muestra grande (y con ayuda de un programa de cómputo), la mediana se sitúa matemáticamente al dividir entre 2 el tamaño muestral y sumar .5. Nótese que este resultado dará la *ubicación ordenada* de la mediana, no la mediana en sí. Ordena las puntuaciones y luego cuenta hasta esta posición. La puntuación X de esta posición es la mediana. Después de hallar la mediana, revisa de nuevo y comprueba si de verdad tu respuesta divide los casos por la mitad. La mediana puede usarse con variables de intervalo/razón, así como con variables ordinales. Finalmente, no confundas la mediana con otro estadístico llamado *rango medio*, que es el punto situado a la mitad entre los valores mínimo y máximo de X .

Cálculo de la mediana (Mdn)

1. Ordena de menor a mayor la distribución de puntuaciones.
2. Ubica la posición de la mediana. Divide entre 2 el tamaño de la muestra, n , para ubicarte cerca de la puntuación que está a la mitad de la distribución. Si n es un número impar, la mediana será un caso real de la muestra; si n es par, la mediana se localizará entre las dos puntuaciones que están a la mitad, y se calculará tomando la media de esas dos puntuaciones. (Matemáticamente, la posición de la mediana se encuentra dividiendo entre 2 el tamaño de la muestra y sumando .5.)

Debilidades potenciales de la mediana: situaciones en las que reportarla sola puede conducir a errores

La mediana se basa en la ubicación ordenada de puntuaciones de una distribución. Es *insensible a valores de puntuaciones* en una desviación, es decir, cualesquiera que sean los valores de X que la rodean, la mediana es la puntuación central determinada por el número de puntuaciones (n) de la muestra. Por ejemplo, las siguientes dos distribuciones de puntuaciones de un examen tienen la misma mediana aun cuando estén compuestas de puntuaciones muy diferentes.

Aula 1: 39, 51, 77, 78, 81



Mdn



Aula 2: 74, 75, 77, 94, 98

Afirmar que la calificación promedio del examen de ambos grupos es 77 sería impreciso porque sugiere que los dos tuvieron igual desempeño. (En realidad, el aula 2 lo hizo mucho mejor, con una *media* de 83.6, comparado con la *media* de 65.2 para el aula 1.) La mediana no se afecta por los valores de X .

Aun cuando es insensible a valores de puntuaciones, la mediana es sensible a (o se ve afectada por) cualquier cambio en el tamaño de la muestra. Por ejemplo, supongamos que en el aula 1 dos estudiantes se presentan tarde al examen y lo hacen mal, lo cual es común en estudiantes que llegan tarde a un examen. Cuando sus puntuaciones se incluyen en la distribución, la mediana cambia de manera drástica de 77 a 51:

Aula 1 (incluye puntuaciones tardías): 34, 36, 39, 51, 77, 78, 81



Mdn

La mediana, entonces, tiene dos debilidades potenciales: (1) es insensible a los valores de las puntuaciones de una distribución y (2) es sensible (o afectada por) cualquier cambio en el tamaño de la muestra. Antes de presentar la mediana, asegúrate que ninguna de estas debilidades potenciales te lleve a conclusiones erróneas.

La moda

La **moda** (Mo) es la puntuación que se presenta con mayor frecuencia en una distribución. Conceptualmente, la moda es la puntuación "más popular". La tabla 4-1 muestra la distribución de edades para una muestra de estudiantes universitarios. La moda es 19 años porque la mayoría de las personas (49 de ellas) tiene esta edad. Observa que la moda es una puntuación X (19 años), *no* una frecuencia, f (49 casos).

TABLA 4-1 Distribución de edades para una muestra de 125 estudiantes universitarios

Especificaciones		Cálculos	
Edad	f	Porcentaje (%)	
18	31	24.8	
Mo → 19	49	39.2	
20	20	16.0	
21	18	14.4	
22	7	5.6	
Total	125	100.0	

La moda Puntuación que ocurre con mayor frecuencia en una distribución.

Cálculo de la moda (Mo)

1. Agrupa las puntuaciones en una distribución de frecuencias.
2. Identifica la moda, que es el valor de X con la mayoría de los casos (es decir, la mayor frecuencia, f).

Es oportuna una nota de precaución. No confundas la moda (la "puntuación que ocurre con mayor frecuencia") con la "mayoría de puntuaciones". Una mayoría simple sería "más de la mitad" o 50 por ciento de los casos de una muestra más uno, por lo menos. Observa que en esta distribución, aunque la puntuación que ocurre más frecuentemente es 19 años, la mayoría de la muestra no tiene 19 años; sólo 39.2 por ciento de la muestra tiene esa edad. Ninguna edad de esta distribución tiene mayoría.

La moda es útil con variables de todos los niveles de medición. La moda es fácil de reconocer en gráficos. En un gráfico de pastel, es la categoría con la rebanada más grande; en un gráfico de barras, es la barra más alta; en un histograma, es la columna más alta; y en un polígono, la puntuación del punto más alto, o pico.

Debilidades potenciales de la moda: situaciones en las que reportarla sola puede conducir a errores

En general, llamada *por sí misma*, la moda es el estadístico de tendencia central menos útil, porque tiene un alcance informativo limitado. Si bien identifica la puntuación que ocurre con más frecuencia, no sugiere nada sobre las puntuaciones que ocurren alrededor de este valor de la puntuación. Así, la moda es muy útil cuando se presenta en conjunción con la mediana y la

FIGURA 4-1
Distribuciones de puntuaciones de varias formas con la misma moda

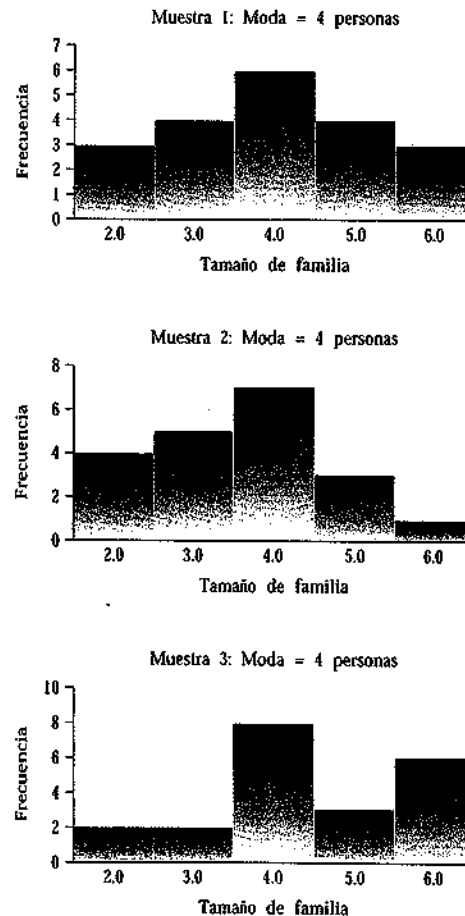


TABLA 4-2 1 Distribución de sueldos en un restaurante de comida rápida

Sueldo \$	f	Clasificación de empleados
5.75	13	Empleados regulares
10.50	2	Gerentes nocturnos
18.90	1	Gerente en Jefe
Total	16	

media. Como veremos más adelante, reportar los tres estadísticos de tendencia central es bastante informativo.

La moda puede ser engañosa cuando se usa sola porque es insensible tanto a los valores de las puntuaciones de una distribución como al tamaño de la muestra. Esto significa que puedes tener cualquier número de distribuciones con formas totalmente diferentes, y aun así todas podrían tener la misma moda, como se ilustra en la figura 4-1. Además, una variable puede tener más de una moda o ninguna moda significativa en absoluto.

Existe al menos una situación en la cual la moda es un estadístico de tendencia central apropiado por sí mismo e informar la media y la mediana es confuso. Esto ocurre cuando las puntuaciones de X son en esencia del mismo valor para todos los casos, excepto para unos cuantos. Un ejemplo es la estructura de sueldos en un restaurante de comida rápida, donde todos, excepto los gerentes, tienen un mismo sueldo bajo. Esta distribución se muestra en la tabla 4-2, donde X es el sueldo por hora y f es la frecuencia de las puntuaciones. La media aquí es \$7.17, y está "inflada" por las puntuaciones extremas de los sueldos de los gerentes. Para alguien que busca empleo, esta media deja la falsa impresión de que el restaurante, en promedio, ofrece un sueldo un tanto arriba del salario mínimo. La mediana es \$5.75, igual que la moda, pero reportar esta mediana lleva a la interpretación incorrecta de que la mitad de los empleados ganan más que esa cantidad, lo cual no es el caso. Informar la moda, \$5.75, significa que a muchos empleados se les paga este sueldo bajo. Ésta es la ilustración más exacta de esta distribución de sueldos.

Estadísticos de tendencia central y el nivel apropiado de medición

Recuerda, como vimos en el capítulo 2, que el nivel de medición de una variable nos dice qué fórmulas matemáticas y estadísticas son apropiadas para dicha variable. La media y la mediana son claramente apropiadas con variables de intervalo/razón. Tiene sentido hablar sobre el peso, la estatura o el ingreso medios. Los estadistas novicios deben evitar el uso de la media y la mediana con variables ordinales. Con variables nominales las medias y las medianas no tienen sentido. La variable nominal *género* es un caso oportuno. Una persona no puede ser un promedio de tanto hombre y tanto mujer; se es o el uno o el otro. Recuerda la tabla 2-5, donde se presenta la distribución de afiliaciones religiosas en la niñez para una muestra de adultos estadounidenses. No tiene sentido preguntar cuál es la media de religión.

Mientras que la media y la mediana se aplican mejor a las variables de intervalo/razón, la moda puede usarse con variables de todos los niveles de medición. De la tabla 2-5 podría-

mos reportar que la moda de religión es "Protestante total" para las principales religiones, "Católico" para cualquier denominación particular, o "Bautista" para cualquier denominación protestante particular.

Curvas de distribución de frecuencias: relaciones entre la media, la mediana y la moda

Puesto que cada uno de los tres estadísticos de tendencia central tiene debilidades potenciales, vale la pena observarlos como un conjunto de estadísticos que se van a interpretar juntos. Estos tres estadísticos son especialmente útiles cuando se examinan de manera gráfica. Una forma imaginativa de entender la relación entre estos tres estadísticos consiste en localizar los valores de cada uno en una curva de distribución de frecuencias.

Una **curva de distribución de frecuencias** es un sustituto de un histograma de frecuencias o polígono donde reemplazamos estos gráficos con una curva suavizada. Esta sustitución es apropiada porque la curva suavizada no se ve tanto como una ilustración de la distribución de la muestra, sino más bien como una estimación de la manera en que se distribuyen las puntuaciones en la población. Al igual que con un histograma, las puntuaciones de una variable se ilustran de izquierda (el más bajo) a derecha (el más alto), es decir, las puntuaciones se ordenan sobre el eje horizontal. El área bajo una curva de distribución de frecuencias representa el número total de sujetos en la población y es igual a una proporción de 1.00 o a un porcentaje de 100 por ciento. Nuestro interés está en evaluar la forma de una distribución y examinar las posiciones relativas de la media, la mediana y la moda, para estimar la forma de una distribución de frecuencias. Las curvas de distribución de frecuencias aplican sólo a niveles de medición de variables de intervalo/razón.

Curva de distribución de frecuencias Es sustituto de un histograma o polígono de frecuencias donde reemplazamos estos gráficos con una curva suavizada. El área bajo la curva representa el número total de sujetos en la población y es igual a una proporción de 1.00 a un porcentaje de 100 por ciento.

La figura 4-2 presenta tres formas muy comunes de curvas de distribución de frecuencias de puntuaciones. Al igual que con nuestros histogramas, el eje horizontal de las curvas representa las puntuaciones de una variable X . El eje vertical (el cual a veces no nos molestamos en dibujar) representa la frecuencia proporcional o frecuencia porcentual; así, la altura de la curva en cualquier valor de X representa la proporción de una muestra o población con esa puntuación.

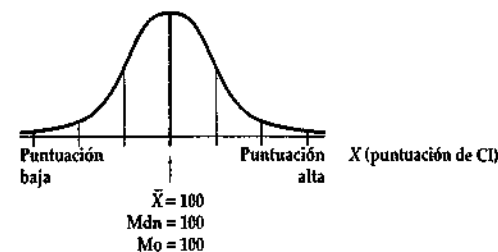
La distribución normal

Una **distribución normal** es aquella donde la media, la mediana y la moda de una variable son iguales entre sí y la distribución de las puntuaciones tiene forma de campana. También nos referimos a esto como una "curva normal". La figura 4-2A ilustra puntuaciones de CI, que están normalmente distribuidos con una media de 100. Una distribución normal es simétrica (es decir, equilibrada en cada lado). Su media, mediana y moda se localizan en el centro de la distribución. La presencia de la mediana aquí asegura la simetría porque, por definición, la mediana divide por la mitad una distribución ordenada de puntuaciones. Puesto que la moda está en el punto central de una distribución normal, el pico de la curva se localiza allí.

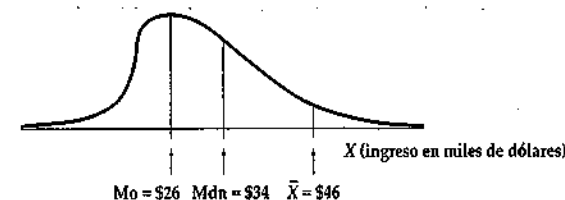
FIGURA 4-2

Curvas de distribución de frecuencias comunes y posiciones relativas de la media, la mediana y la moda, donde X es una variable de intervalo/razón (datos ficticios)

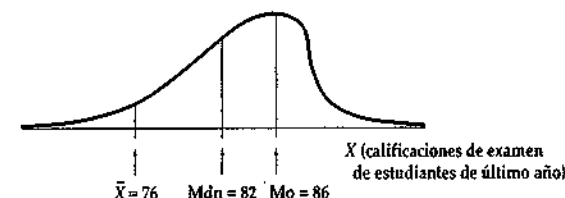
A. Distribución normal o curva normal



B. Distribución sesgada positivamente o sesgada a la derecha



C. Distribución sesgada negativamente o sesgada a la izquierda



Distribución normal Curva de distribución de frecuencias donde la media, la mediana y la moda de una variable son iguales entre sí y la distribución de las puntuaciones tiene forma de campana.

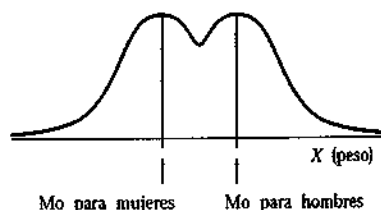
Distribuciones sesgadas

Una **distribución sesgada** es aquella en la cual la media, la mediana y la moda de una variable son desiguales y algunos de los sujetos tienen puntuaciones sumamente altas o bajas. Cuando éste es el caso, la distribución se alarga hacia un lado, como la hoja de una espada o de una brocheta (*skewer*), de ahí el nombre de *sesgada* (*skewed*) (figura 4-2B y C).

Distribución sesgada Curva de distribución de frecuencias en la cual la media, la mediana y la moda de una variable son desiguales y algunos de los sujetos tienen puntuaciones sumamente altas o bajas.

FIGURA 4-3

Distribución bimodal de pesos de hombres y mujeres



Las posiciones de la media, la mediana y la moda son predecibles para las curvas de distribución sesgadas. Un **sesgo a la derecha (o positivo)** tiene puntuaciones extremas en el extremo positivo de la distribución de puntuaciones (figura 4-2B). Por ejemplo, el ingreso familiar en Estados Unidos está sesgado positivamente; la mayoría de las familias ganan bastante dinero, pero pocas son sumamente ricas. Las puntuaciones extremas altas inflan la media, "jalándola" en dirección positiva. La moda es la medida de tendencia central con la menor puntuación calculada. La mediana será igual a la media o a la moda o, más probablemente, caerá entre éstas.

El **sesgo a la izquierda (o negativo)** tiene puntuaciones extremas en el final bajo o negativo de la distribución de puntuaciones (figura 4-2C). Por ejemplo, las puntuaciones del examen en un curso del último año en la universidad tienden a estar sesgadas a la izquierda. La mayoría de los estudiantes de último año obtiene altas puntuaciones, pero pocos se quedan en la dirección negativa. Estas pocas puntuaciones extremas bajas desinflan la media, jalándola en la dirección negativa. La moda es la mayor puntuación calculada, y la mediana cae entre la media y la moda.

Ya sea con un sesgo a la izquierda o la derecha, si la mediana no cae entre la media y la moda esto sugiere que la distribución está singularmente formada. Una distribución así es una distribución bimodal, la cual tiene dos modas o picos. Por ejemplo, la variable peso para una muestra que incluye a hombres y mujeres produciría una distribución bimodal, con la moda más alta que resulta del hecho de que en promedio los hombres son más pesados que las mujeres (figura 4-3).

Uso de los datos de una muestra para estimar la forma de una distribución de puntuaciones en una población

Ya sea con variables intervalo/razón cuando al calcular los estadísticos de tendencia central e histogramas para datos de una muestra, los datos para una variable con frecuencia aparecen ligeramente sesgados. Esto no garantiza, sin embargo, que las puntuaciones de la variable estén sesgadas en la población de la que se tomó la muestra. El sesgo en los datos de la muestra puede deberse al error muestral. En otras palabras, una segunda muestra de la población parecería normal o ligeramente sesgada en la otra dirección.

Los estadísticos de sesgo se emplean para determinar si los datos de la muestra están tan sesgados que sugieren que las puntuaciones de la población están sesgadas. No vamos a calcular un estadístico de sesgo a mano. Los programas de cómputo, sin embargo, proporcionan estadísticos de sesgo, y uno común está disponible con las aplicaciones de cómputo opcionales que acompañan este texto. Cuando el valor absoluto de este estadístico de sesgo (su valor ignorando el signo de más y de menos) es mayor a 1.2, la distribución podría estar significativamente sesgada, dependiendo de la forma de la distribución, así como del tamaño de la muestra. Unos pocos valores extremos de una muestra grande tendrán poco efecto en los

estadísticos. Si el valor absoluto de este estadístico de sesgo es mayor que 1.6, sin embargo, sin importar el tamaño de la muestra, la distribución probablemente esté sesgada; entonces reportar la media de X de la muestra como un estimado de la media de la población sería engañoso, a causa de la distorsión potencial de la media por las puntuaciones extremas. Aparte de la cuestión de describir con precisión la forma de una distribución, el sesgo es una preocupación con la estadística inferencial. Como veremos en capítulos posteriores, al probar una hipótesis sobre la relación entre dos variables, una variable sesgada exige trabajo adicional para evitar conclusiones incorrectas. Se identificarán tales casos conforme se encuentren.

Como veremos en el capítulo 5, cuando una distribución no esté sesgada o de otra manera tenga una forma particularmente extraña, la media es el estadístico de tendencia central a elegir. Esto es especialmente válido para reportes dirigidos al público en general, cuyos miembros pueden sentirse abrumados con más de un estadístico. Sin embargo, si una distribución está sesgada, la mediana es el estadístico que deba reportarse. La mediana minimiza el error al describir una distribución sesgada, porque cae entre la media y la moda, como se ilustra en la figura 4-2B y C. Como la más central de los tres estadísticos, la mediana es el mejor de las tres pobres opciones para una distribución sesgada, cuando sólo un estadístico debe reportarse.

Para audiencias científicas, las distribuciones sesgadas se registran informando los tres estadísticos de tendencia central y quizás incluyendo un gráfico para transmitir con precisión la forma de la distribución. A veces una distribución sesgada es muy informativa. Por ejemplo, las estancias en el hospital están positivamente sesgadas. En un año dado, la mayoría

TABLA 4-3 | Características, aspectos fuertes, y debilidades potenciales de la media, mediana y moda

Estadístico de tendencia central	Definición	Aspectos fuertes y aplicaciones		
		Niveles apropiados de medida	Aplicación a formas de distribución de puntuaciones	Debilidades potenciales
Media	Valor de X si todas las puntuaciones son iguales	Intervalo/razón	Abierta a operaciones matemáticas cuando una distribución tiene forma normal	Su cálculo es distorsionado por valores extremos o un sesgo en la curva de la distribución
Mediana	Puntuación en la mitad de una distribución ordenada; puntuación por arriba de la cual queda la mitad de las puntuaciones y, por debajo, la otra mitad	Intervalo/razón	Preferible cuando la distribución es sesgada	Insensible a los valores de X de la distribución, pero sensible a cambios en el tamaño muestral
Moda	Puntuación que ocurre con más frecuencia en una distribución; la puntuación "más popular"	Nominal, ordinal, intervalo/razón	Preferida cuando prácticamente todas las puntuaciones (o categorías) de la distribución son iguales	Insensible a los valores de X e insensible a cómo se distribuyen las puntuaciones alrededor de X

de las personas no pasan algún día o pasan muy pocos en el hospital. Pero un porcentaje sustancial pasa mucho tiempo, y unos pocos "se sesgan" al permanecer semanas o meses en el hospital. Tal sesgo estimula la reflexión sobre los predictores de estancias largas. ¿Puedes pensar en hipótesis que expliquen el sesgo de estancias en el hospital?

Como veremos en el capítulo 5, en general la media es el estadístico de tendencia central más valioso porque permite mayor flexibilidad en los cálculos matemáticos. Casi siempre, la mediana y la moda representan callejones sin salida porque no ofrecen operaciones matemáticas adicionales que valgan la pena. Se gana poco con informar de ellas. Siempre que sea posible, la media es la medición sumaria que debe usarse, en especial con estadísticos inferenciales. Debido a esto, con frecuencia ajustamos distribuciones sesgadas para "hacerlas normales", de manera que podamos usar la media. Más adelante en este texto se discuten las especificaciones de este tipo de control del error. La tabla 4-3 resume las propiedades de los tres estadísticos de tendencia central.

Organización de los datos para calcular los estadísticos de tendencia central

Existen dos formatos comunes para organizar los datos y calcular los estadísticos de tendencia central en tales datos. Un formato es una hoja de cálculo que por lo general se usa para la entrada de datos para computadora, pero las hojas de cálculo también se usan en negocios, gobierno y grupos comunitarios para conservar registros de la organización. Programas computarizados de hojas de cálculo, como *Lotus 1-2-3*, *Excel*, y *Corel Quattro Pro*, están especialmente diseñados para este fin. Estos formatos de hoja de cálculo evolucionaron a partir de la manera lógica de resolver problemas manualmente, con sólo poner en lista las puntuaciones de una variable en una columna vertical.

El segundo formato común para realizar cálculos es un formato de distribución de frecuencias. En éste, las puntuaciones de una variable se escriben en una columna y la frecuencia de cada puntuación en otra (como las distribuciones de frecuencias del capítulo 2). Este formato es típico de una salida de resultados de una computadora. Ahora resolvamos un problema sencillo para ilustrar el uso de ambos formatos.

Formato de hoja de cálculo para calcular estadísticos de tendencia central

Supongamos que nos interesa saber con qué frecuencia es que los estudiantes de cinematografía, de un departamento de comunicaciones de la universidad, estudian su disciplina asistiendo a cines con películas de estreno. Recolectamos una muestra aleatoria de 19 estudiantes y pedimos a cada uno mencionar las nuevas películas que vieron en el cine el mes pasado, y registramos los siguientes resultados: 2, 6, 4, 5, 2, 3, 4, 3, 6, 4, 3, 3, 5, 4, 5, 2, 3, 4, 3. La tabla 4-4 presenta estos datos en un formato de hoja de cálculo con los cálculos necesarios para calcular la media. Las puntuaciones se ordenan para facilitar el cálculo de la mediana y la moda.

Primero, calculemos la media:

$$\bar{X} = \frac{\sum X}{n} = \frac{72}{19} = 3.79 \text{ películas}$$

En segundo término, calculemos la mediana. Ya ordenamos las puntuaciones, lo cual es necesario para calcular la mediana. El tamaño muestral ($n = 19$) dividido entre 2 es alrededor de nueve casos, y como n es impar, determinamos que el décimo caso es la mediana. En la

TABLA 4-4 Datos organizados en hoja de cálculo: número de películas de estreno vistas en el último mes (X)

Especificaciones		
Número del sujeto	Iniciales del sujeto	X
1	BH	2
2	KP	2
3	JN	2
4	TW	3
5	JD	3
6	WA	3
7	KM	3
8	BC	3
9	CR	4
10	ML	4
11	MW	4
12	MF	4
13	JS	4
14	BY	4
15	LL	5
16	WF	5
17	CM	5
18	BL	6
19	SH	6
$n = 19$		$\sum X = 72$ películas

hoja de cálculo contamos hacia abajo hasta el décimo caso y descubrimos que la mediana son cuatro películas:

$Mdn = 4$ películas

Por último, calculamos la moda. La observación de los datos ordenados en la tabla 4-4 revela que la puntuación que ocurre con más frecuencia es 4:

$Mo = 4$ películas

Obviamente, el empleo de una hoja de trabajo para hacer cálculos manualmente con un gran número de casos sería difícil. El propósito de un ejercicio hecho a papel y lápiz es entender las características fundamentales de la estadística. En un trabajo real de investigación, se emplean paquetes de software de estadística y hojas de cálculo computarizadas para ahorrar tiempo y reducir errores en cálculos.

Formato de distribución de frecuencias para calcular la moda

La tabla 4-5 presenta los mismos datos sobre los 19 estudiantes de cinematografía, pero emplea un formato de distribución de frecuencias. Trabajando a partir de la hoja de cálculo de

TABLA 4-5 1 Datos organizados en un formato de distribución de frecuencias: número de películas de estreno vistas el último mes (X)

Especificaciones	
X	f
2	3
3	5
4	6
5	3
6	2
$n = 19$	

la tabla 4-4 (como lo haría una computadora), en la tabla 4-5 vemos que hay una frecuencia de tres estudiantes que reportan dos películas, cinco que reportan tres películas, y así sucesivamente.

El cálculo de la moda es muy fácil con el formato de distribución de frecuencias. En la tabla 4-5 simplemente observamos la columna que indica las frecuencias (es decir, la columna f) y vemos cuál puntuación se presentó con más frecuencia. Más estudiantes (seis de ellos) vieron cuatro películas que ningún otro número de películas para el mes:

$$Mo = 4 \text{ películas}$$

La salida de distribución de frecuencias y los estadísticos descriptivos básicos son características estándar de los paquetes de software de estadística.

Insensatez y falacias estadísticas: mezcla de subgrupos en el cálculo de la media

Debido a que la media es susceptible de distorsión por valores y puntuaciones extremos, debemos describir con claridad qué casos o qué sujetos se incluyen en su cálculo. Organizaciones tales como son por ejemplo empresas o instituciones escolares, intencionalmente o no, por lo general reportan medias que son irreales. Por ejemplo, el vocero de un distrito escolar público puede reportar que el sueldo medio de sus maestros es \$45 000. Cuando esto ocurra, es probable que los maestros se reúnan en el aula de descanso de la facultad y se pregunten entre sí: ¿Quién entre nosotros gana tanto dinero? Por supuesto, los maestros no son tontos. Ellos saben de inmediato que quien realizó los cálculos "mezcló los rangos de estatus" e incluyó al personal de mayor salario, por ejemplo, consejeros académicos, auxiliares de los directores, y directores, todos ellos certificados como docentes pero rara vez dan clases. Estos administradores quizá hayan sido incluidos porque el "estadista" simplemente pidió a la computadora calcular el sueldo medio para todos los maestros certificados sin tener en cuenta el rango. Cuando se incluyó este personal bien pagado, sus altos salarios sesgaron la media. Para evitar tal insensatez estadística, deben informarse por separado las medias para subgrupos distintos.

La mezcla de rangos de estatus resulta a veces en una media que no se ajusta a ningún grupo. Por ejemplo, una compañía puede tener sólo dos rangos de empleados: obreros que

promedian cerca de 30 000 dólares al año, y gerentes que promedian cerca de 70 000 dólares al año. Si estos dos grupos son aproximadamente del mismo tamaño, el sueldo medio para la compañía entera estaría cercano a \$50 000. Curiosamente, ningún empleado de la compañía gana un sueldo cercano a esa cantidad.

Otro ejemplo es la edad media de asistentes de una clase nocturna de tercer grado de una escuela primaria. La edad media se calculará en 20 años más menos, aunque todos ahí tendrán 8 o 9 años (los niños) o alrededor de 30 (los padres). La media es ciertamente impropia para resumir esta distribución de edades.

RESUMEN

1. Un estadístico de tendencia central es el que proporciona una estimación típica, usual, normal o promedio que se encuentre en una distribución de puntuaciones brutas.
2. Hay tres medidas de tendencia central: la media, la mediana y la moda. Cada una de ellas tiene puntos fuertes y débiles.
3. Los valores relativos de las tres estadísticas de tendencia central nos informan acerca de la forma de una distribución de puntuaciones.
4. La media y la mediana son apropiadas con variables de intervalo/razón. Con variables nominales, las medias y las medianas no tienen sentido. La moda se puede usar con variables de todos los niveles de medida.
5. La media es el estadístico de tendencia central más útil.
6. El cálculo de la media resulta afectado por valores extremos y por un sesgo en la distribución de las puntuaciones.
7. La mediana es una puntuación posicional, la puntuación central en una distribución ordenada. Es igual al 50o. percentil.
8. Cuando una distribución está sesgada, la mediana es el estadístico a elegir porque su valor caerá entre la media y la moda y, así, minimiza el error.
9. Recuerda ordenar las puntuaciones de menor a mayor antes de calcular la mediana.
10. La mediana es insensible a los valores de las puntuaciones de una distribución. La mediana es sensible a un cambio en el tamaño muestral.
11. La moda es la puntuación o categoría que ocurre con más frecuencia en una distribución. La moda puede verse como la puntuación o categoría más popular, pero no debe confundirse la moda con "la mayoría de puntuaciones".
12. La moda es fácil de ubicar en tablas y gráficos. Al identificar la moda, debes tener cuidado en recordar que es una puntuación (X), no una frecuencia (f).
13. La moda es la menos útil de las medidas de tendencia central por sí misma, porque tiene un alcance informativo limitado, esto es, nos dice poco. La moda es insensible a los valores de puntuaciones de una distribución e insensible al tamaño muestral. Dos distribuciones de puntuaciones pueden tener formas radicalmente diferentes, pero tener la misma moda.

14. Una curva de distribución de frecuencias es un sustituto para un histograma o polígono de frecuencias en donde sustituimos estos gráficos con una curva suavizada.
15. Las ubicaciones relativas de la media, mediana y moda sobre el eje X son predecibles para ciertas formas de curvas de distribución. En una distribución normal o "curva normal", la media, mediana y moda de la variable son iguales. En una distribución sesgada negativamente, la media tendrá el valor de X más bajo, la moda más alta, y la mediana caerá entre ellas. En una distribución sesgada positivamente, la media tendrá el valor de X más alto, la moda el más bajo y la mediana caerá entre ellos.
16. Por lo general, la media de dos grupos combinados *no* es simplemente la suma de las medias dividida entre 2. Esto sólo funciona cuando los dos grupos son del mismo tamaño muestral.

EXTENSIONES DEL CAPÍTULO EN EL SITIO WEB THE STATISTICAL IMAGINATION

Las extensiones del capítulo 4 del material de texto disponibles en el sitio web *The Statistical Imagination*, en www.mhhe.com/ritchey2, incluyen ilustraciones donde la media y la mediana se pueden usar con variables ordinales ordenadas con ciertas características.

FÓRMULAS PARA EL CAPÍTULO 4

Cálculo de la media:

Trabajando con una hoja de cálculo:

$$\bar{X} = \frac{\sum X}{n}$$

Cálculo de la media combinada de dos grupos (a partir de puntuaciones individuales):

$$\bar{X}_{(\text{grupos 1 y 2 combinados})} = \frac{\sum X_{(\text{grupo 1})} + \sum X_{(\text{grupo 2})}}{n_{(\text{grupo 1})} + n_{(\text{grupo 2})}}$$

Cálculo de la media combinada de dos grupos (a partir de medias de grupo):

$$\text{Como } \bar{X} = \frac{\sum X}{n}, \sum X = (n)(\bar{X})$$

Sustituye para obtener:

$$\bar{X}_{(\text{grupos 1 y 2 combinados})} = \frac{n_{(\text{grupo 1})}\bar{X}_{(\text{grupo 1})} + n_{(\text{grupo 2})}\bar{X}_{(\text{grupo 2})}}{n_{(\text{grupo 1})} + n_{(\text{grupo 2})}}$$

Cálculo de la mediana:

1. Ordena la distribución de puntuaciones de la más baja a la más alta.
2. Ubica la posición de la mediana. Divide entre 2 el tamaño muestral, n , para obtener la puntuación central de la distribución. Si n es un número impar, la mediana será un caso real en la muestra. Si n es un número par, la mediana estará entre las dos puntuaciones

centrales y se calcula tomando la media de estas dos puntuaciones. (Matemáticamente, la posición de la mediana se encuentra dividiendo entre 2 el tamaño de la muestra y sumando .5.)

Cálculo de la moda:

1. Agrupa las puntuaciones en una hoja de cálculo de puntuaciones ordenadas sin elaborar o formato de distribución de frecuencias.
2. Identifica M_o = valor de X con la frecuencia mayor.

PREGUNTAS PARA EL CAPÍTULO 4

1. Para cada estadístico de tendencia central, ¿variables de qué niveles de medición son apropiadas?
2. Define la media, la mediana y la moda. Especifica las limitaciones potenciales de cada una.
3. ¿Por qué es mejor calcular las tres medidas: la media, mediana y la moda, que confiar en una de ellas?
4. Como regla general, es incorrecto calcular la media para dos grupos combinados dividiendo simplemente entre 2 la suma de sus medias separadas. ¿Cuál es la excepción a esta regla?
5. Si una distribución de puntuaciones está sesgada, ¿qué único estadístico de tendencia central es más apropiado para una audiencia pública? ¿Por qué?
6. En general, la moda de una distribución es el estadístico de tendencia central menos útil. ¿Bajo qué circunstancias, no obstante, es el estadístico de tendencia central más apropiado a reportar?
7. Si la edad modal de una distribución es 22 años, ¿significa esto que una mayoría de las personas de esta población tiene 22 años de edad? Explica.
8. ¿Cómo se ubica la moda en un histograma, un polígono y una curva de distribución de frecuencias?
9. En una curva de distribución de frecuencias, ¿qué representan los ejes horizontal y vertical?
10. Describe las características de una curva de distribución de frecuencias normal.
11. Expresa en términos generales cómo un sesgo a la izquierda, en una distribución de frecuencias, afecta a los tres promedios comunes: media, mediana y moda.
12. Expresa en términos generales cómo un sesgo a la derecha, en una distribución de frecuencias, afecta a los tres promedios comunes: media, mediana y moda.
13. Supongamos que una distribución de edades tiene una media de 55 años, una moda de 28 años y una mediana de 34 años. ¿Cuál es la forma probable de la curva de distribución de frecuencias de esta variable?
14. Como se ilustra en "Insensatez y falacias estadísticas" de este capítulo, la media de una variable puede ser una mala medida de tendencia central cuando se mezclan los rangos de estatus dentro de una población. Da un ejemplo de cómo mezclar rangos puede resultar en una media que no se ajusta en absoluto a ningún rango.

EJERCICIOS PARA EL CAPÍTULO 4

Conjunto de problemas 4A

Recuerda incluir las fórmulas, estipular las unidades de medida y contestar la pregunta.

- 4A-1. Dados los datos siguientes con X = edad, calcula la edad modal, edad mediana y edad media. Empieza por organizar los datos en una tabla de hoja de cálculo con las puntuaciones ordenadas bajo una columna marcada " X (ordenados)".

X	X (cont.)
14	14
15	17
19	19
19	22
22	28

- 4A-2. Siete trabajadores de oficinas entraron a un concurso de reducción de peso. Tras unas semanas de someterse a dieta, los pesos que bajaron (en libras) fueron como sigue: 5, 7, 3, 0, 2, 4 y 3. Calcula la pérdida de peso mediana y media. Los datos de X = libras perdidas.
- 4A-3. Los expertos en demografía estudian las poblaciones de varios estados. Un sujeto de interés es el crecimiento o disminución del tamaño de una población, que es afectado por el índice de natalidad, longevidad (tiempo que viven las personas con base en edades en que normalmente mueren), y cuántos se establecen en un lugar o se van de éste (emigración). Una variable es la edad de mortalidad (es decir, la edad a su fallecimiento). Supongamos que en la nación A, la edad de mortalidad modal es 55, la mediana es 60 y la media es 65. En la nación B, la media también es 65, pero la moda es 75 y la mediana es 70.
- Con esta información, construye las curvas de frecuencias de edad de mortalidad para cada nación.
 - ¿Qué nación parece encontrarse mejor en términos de longevidad?
- 4A-4. Los cinco miembros de una familia trabajan. Sus salarios por hora son: \$30, \$10.50, \$5.15, \$12, y \$6. Los datos de X = salario por hora.
- Calcula la media y la mediana.
 - En comparación con las otras puntuaciones, ¿cómo llamaríamos al salario de \$30 por hora?
 - ¿Cuál es su efecto en el cálculo de la media?
 - Ajusta esta peculiaridad, recalculando la media sin ella.
- 4A-5. Los siguientes son precios de colegiaturas anuales para cinco universidades norteamericanas importantes: \$10 000, \$29 000, \$8 000, \$12 500, \$11 300. Los datos Y = precio de colegiatura.
- Calcula la media y la mediana.
 - En comparación con las otras puntuaciones, ¿cómo llamaríamos al precio de colegiatura de \$29 000?

- ¿Cuál es su efecto en el cálculo de la media?

- Ajusta esta peculiaridad, recalculando la media sin ella.

- 4A-6. La edad media de los 47 hombres del club de bridge de Sparkesville es 54.8 años. La edad media de las 62 mujeres del club es 56.4 años. ¿Cuál es la edad media de los 109 miembros? Los datos X = edad.
- 4A-7. En un experimento para ver si los pollos pueden distinguir colores, le dan premios de granos de maíz a un pollo cuando pica correctamente un cojincillo de igual color. Los tiempos de reacción se miden al centésimo de segundo más cercano. Los tiempos de reacción de Flossy son como sigue: 1.32, 1.45, 1.21, 1.05, .97, .91, .93, .93, .96, .93, .88, .94, .98.

Los datos X = tiempo de reacción.

- Organiza los datos en una tabla de hoja de cálculo con las puntuaciones ordenadas bajo una columna marcada " X (ordenada)".
 - Calcula los tiempos de reacción de media, mediana y modal de Flossy.
 - Describe la forma de la distribución de los tiempos de reacción de Flossy.
- 4A-8. Dados los siguientes estadísticos y lo que sabemos acerca de cómo están relacionadas dentro de una distribución de puntuaciones, describe la probable forma de la distribución para cada una de las variables citadas. Traza la curva indicando las ubicaciones relativas de la media, mediana y moda.

Variable	$\bar{X}$	Mdn	Mo	Forma de curva	Traza de curva
Edad (años)	30	35	39		
Tamaño de la familia	4.1	3.0	2.0		
Años empleado	11	8	7		
Peso (libras)	160	132	134		

Conjunto de problemas 4B

Recuerda incluir fórmulas, estipular unidades de medida y contestar la pregunta.

- 4B-1. Los datos siguientes son para la variable Y = distancia desde el lugar de trabajo (en millas) para los empleados de un vendedor de copiadoras. Calcula las puntuaciones de la media y la mediana. Empieza por organizar los datos en una tabla de hoja de cálculo con puntuaciones ordenadas bajo una columna marcada " Y (ordenados)".

Y	Y (cont.)
13	10
9	11
6	14
3	5
12	7

- 4B-2. Las puntuaciones de la parte analítica del examen de registros de graduados (GRE) de cinco candidatos a un programa de graduados fueron como sigue: 700, 625, 640,

590 y 600. Calcula las puntuaciones de la media y la mediana. Los datos X = puntuación GRE.

- 4B-3. Al evaluar los porcentajes de delincuencia entre dos ciudades, un criminalista calcula X = número promedio de vehículos robados por día (en un periodo de seis meses). Para la ciudad A, la moda de X es 15 vehículos, la mediana es 20 y la media es 25. Para la ciudad B, la media es también 25, pero la moda es 35 y la mediana es 30.

- A partir de esta información, construye las curvas de frecuencias para cada una de las ciudades.
- ¿En cuál ciudad te sentirías más seguro de estacionar tu auto en la calle? ¿Por qué?

- 4B-4. Los siguientes son promedios puntuales de calificaciones (GPA en una escala de 4 puntos) de estudiantes en un programa de clases prácticas: 1.7, 2.6, 2.3, 3.9, 2.2, 1.9, 2.1. Los datos Y = GPA.

- Organiza los datos en una tabla de hoja de cálculo con las puntuaciones ordenadas bajo una columna marcada " X (ordenada)".
- Calcula la media y la mediana de Y .
- En comparación con las otras puntuaciones, ¿cómo llamaríamos al GPA de 3.9?
- ¿Cuál es su efecto en el cálculo de la media?
- Ajusta esta peculiaridad, recalculando la media sin ella.

- 4B-5. Las siguientes son calificaciones de examen final para cinco estudiantes, pasantes de ciencias sociales, en una importante universidad urbana: 90, 88, 64, 92, 87. Los datos X = calificación de examen.

- Calcula la media y mediana de calificación del examen.
- En comparación con las otras puntuaciones, ¿cómo llamaríamos a la puntuación de 64?
- ¿Cuál es su efecto en el cálculo de la media?
- Ajusta esta peculiaridad, recalculando la media sin ella.

- 4B-6. Supongamos que las siguientes son las edades medias de pacientes adictos a sustancias, en un hospital local para su tratamiento, separadas por tipo de adicción. Calcula la edad media de todos los pacientes adictos a sustancias del hospital. Los datos X = edad.

	Adicción primaria			
	Cocaína ($n = 44$)	Cocaína crack ($n = 29$)	Heroína ($n = 24$)	Alcohol ($n = 69$)
Edad media (años)	29.8	23.4	34.6	42.9

- 4B-7. Los promedios de bateo de la alineación inicial del equipo de ligas pequeñas, los *Dodgers Bola Rápida*, son como sigue: .360, .200, .350, .355, .230, .345, .360, .380, y .400. Los datos X = promedio individual de bateo.

- Organiza los datos en una tabla de hoja de cálculo con las puntuaciones ordenadas.

- Calcula los porcentajes de bateo de media, mediana y moda del equipo.
- Describe la forma de la distribución.

- 4B-8. Dados los siguientes estadísticos y lo que sabemos acerca de cómo están relacionadas dentro de una distribución de puntuaciones, describe la probable forma de la distribución para cada una de las variables citadas. Traza la curva indicando las ubicaciones relativas de la media, mediana y moda.

Variable	$\bar{X}$	Mdn	Mo	Forma de la curva	Traza de la curva
Estatura (pulgadas)	70	68	66		
Exámenes este semestre	10	13	15		
Puntuación de espiritualidad	30	30	30		
Presupuesto de abarrotos	\$130	\$109	\$104		

Conjunto de problemas 4C

Recuerda incluir las fórmulas, estipular las unidades de medida y contestar la pregunta.

- 4C-1. A partir de la siguiente serie de mediciones de estaturas (en pulgadas), calcula las estaturas modal, mediana y media. Los datos X = estatura (en pulgadas). Empieza por organizar los datos en una tabla de hoja de cálculo con las puntuaciones ordenadas bajo una columna marcada " X (ordenada)".

X	X (cont.)
60	78
70	59
68	67
72	74
69	70

- 4C-2. Supongamos que los siguientes son números de victorias de conferencia entre siete equipos colegiales de baloncesto: 12, 8, 7, 9, 11, 5 y 4. Organiza los datos en una tabla de hoja de cálculo con las puntuaciones ordenadas. Calcula el número de mediana y media de victorias entre estos equipos. Los datos X = número de victorias o "ganés".
- 4C-3. Al comparar las puntuaciones de un examen entre los estudiantes, el director de un departamento calcula las calificaciones del examen de dos grupos. Para el grupo A, la moda es 75 puntos, la mediana es 80 y la media es 85. Para el grupo B, la media es también 85, pero la moda es 95 y la mediana es 90. Los datos X = calificación de examen.

- A partir de esta información, construye las curvas de frecuencias de las calificaciones del examen para cada grupo.
- ¿Cuál grupo parece estar mejor en este examen en particular?

- 4C-4. Los siguientes son salarios anuales (Y) entre siete médicos empleados en una zona urbana importante: \$88 000, \$94 000, \$86 000, \$110 000, \$212 000, \$115 000 y \$97 000.

- Organiza los datos en una tabla de hoja de cálculo con las puntuaciones ordenadas bajo una columna marcada "Y (ordenada)".
- Calcula el salario medio y mediano.
- En comparación con otras puntuaciones, ¿cómo llamaríamos al salario de \$212 000?
- ¿Cuál es su efecto en el cálculo de la media?
- Ajusta esta peculiaridad, recalculando la media sin ella.

4C-5. Las siguientes son evaluaciones del rendimiento de empleados (completadas por sus supervisores) de una importante empresa fabricante de software. Cada empleado evaluado en una escala de 0 a 10, con base en su rendimiento de varios indicadores establecidos. Los datos Y = evaluación de empleado.

- Organiza los datos en una tabla de hoja de cálculo con las puntuaciones ordenadas bajo una columna marcada "Y (ordenada)".
- Calcula la media y la mediana.
- En comparación con otras puntuaciones, ¿cómo llamaríamos a la evaluación de 3?
- ¿Cuál es su efecto en el cálculo de la media?
- Ajusta esta peculiaridad, recalculando la media sin ella.

Y
8
3
9
7
8
6
9
8
7

4C-6. Supongamos que el peso medio de 34 hombres que participan en un programa local de reducción de peso es 228 libras, en tanto que el peso medio de 46 mujeres que participan en el mismo programa es de 194 libras. ¿Cuál es el peso medio de los 80 participantes? Los datos X = peso (en libras).

4C-7. Supongamos que nueve amigos están compitiendo todos contra todos en una carrera cronometrada de 40 yardas. El tiempo de cada uno de los participantes (en segundos) es como sigue: 4.8, 5.2, 4.7, 4.9, 5.4, 4.8, 4.9, 4.8 y 5.3. Los datos X = tiempo (en segundos).

- Organiza los datos en una tabla de hoja de cálculo con las puntuaciones ordenadas.
- Calcula los tiempos de media, mediana y moda de la carrera de 40 yardas del grupo.
- Describe la forma de la distribución.

4C-8. Dadas las siguientes estadísticas y lo que sabemos acerca de cómo están relacionadas dentro de una distribución de puntuaciones, describe la probable forma de la distribución para cada una de las variables citadas. Traza la curva indicando las ubicaciones relativas de la media, mediana y moda.

Variable	$\bar{X}$	Mdn	Mo	Forma de la curva	Traza de la curva
Peso (libras)	195	205	215		
Dinero para gastar	\$150	\$140	\$125		
Puntuación escala de depresión	25	25	25		
Edad (años)	30	40	50		

Conjunto de problemas 4D

Recuerda incluir fórmulas, estipular unidades de medida y contestar la pregunta.

4D-1. Dadas las siguientes mediciones de peso de varios amigos, calcula los pesos modal, mediano y medio. Los datos X = peso (en libras). Empieza por organizar los datos en una tabla de hoja de cálculo con las puntuaciones ordenadas bajo una columna marcada "X (ordenada)".

X	X (cont.)
158	180
180	195
169	200
190	195
180	160

4D-2. Bajo la supervisión de sus maestros, un pequeño grupo de estudiantes adolescentes decidió evaluar su crecimiento general en estatura (en pulgadas), para un periodo de 18 meses. Las diferencias en sus estaturas entre el inicio (tiempo 1) y terminación (tiempo 2) son como sigue: 4.4, 6.0, 3.6, 2.9, 4.3, 3.6, 2.9, 4.2 y 2.8. Calcula el crecimiento mediano y medio. Los datos X = crecimiento en 18 meses (en pulgadas).

4D-3. Un investigador está interesado en comparar patrones de ingreso familiar en dólares de Estados Unidos (X) entre dos comunidades de clase media alta. Para la Comunidad 1, el ingreso modal familiar es \$80 000, la mediana es \$90 000 y la media es \$100 000. Para la Comunidad 2, la media es también \$100 000, pero la moda es \$120 000 y el ingreso mediano familiar es \$110 000.

- De esta información, construye las curvas de frecuencias de ingreso familiar para cada comunidad.
- ¿Qué comunidad parece ser más acaudalada con respecto al ingreso familiar?

4D-4. Los siguientes son tamaños de grupo para siete grupos de introducción a la sociología en una importante universidad urbana: 65, 79, 72, 115, 84, 87 y 78. Los datos Y = número de estudiantes.

- a) Organiza los datos en una tabla de hoja de cálculo con las puntuaciones ordenadas bajo una columna marcada "Y (ordenada)".
- b) Calcula la media y la mediana.
- c) En comparación con otras puntuaciones, ¿cómo llamaríamos al tamaño del grupo de 115?
- d) ¿Cuál es su efecto en el cálculo de la media?
- e) Ajusta esta peculiaridad, recalculando la media sin ella.
- 4D-5.** Los siguientes son números de empleados pagados entre nueve subsidiarias de una importante empresa financiera internacional. Los datos X = número de empleados.
- a) Organiza los datos en una tabla de hoja de cálculo con las puntuaciones ordenadas bajo una columna marcada "Y (ordenada)".
- b) Calcula la media y la mediana.
- c) En comparación con otras puntuaciones, ¿cómo llamaríamos a la subsidiaria con sólo 67 empleados?
- d) ¿Cuál es su efecto en el cálculo de la media?
- e) Ajusta esta peculiaridad, recalculando la media sin ella.

X
212
198
283
176
191
254
67
187
193

- 4D-6.** La puntuación media de examen de registro de graduados (GRE) de los 39 solicitantes hombres al departamento de sociología de la universidad central es 1140 puntos GRE. La puntuación media para las 54 solicitantes mujeres es 1210. ¿Cuál es la puntuación media GRE para los 93 solicitantes? Los datos X = puntuación GRE.
- 4D-7.** Nueve amigos compiten todos contra todos en una liga de fútbol americano. Las yardas de pase para los mariscales de campo estrellas de cada competidor de la semana anterior son como sigue: 283, 205, 183, 197, 296, 315, 304, 227 y 296. Haz X = yardas de pase.
- a) Organiza los datos en una tabla de hoja de cálculo con las puntuaciones ordenadas bajo una columna marcada "Y (ordenada)".
- b) Calcula las yardas de pase media, mediana y moda.
- c) Describe la forma de la distribución.

- 4D-8.** Dados los siguientes estadísticos y lo que sabemos acerca de cómo están relacionadas dentro de una distribución de puntuaciones, describe la probable forma de la distribución para cada una de las variables citadas. Traza la curva indicando las ubicaciones relativas de la media, mediana y moda.

Variable	$\bar{X}$	Mdn	Mo	Forma de la curva	Traza de la curva
Gastos de entretenimiento	\$163	\$154	\$139		
Puntos de la escala de religiosidad	30	30	30		
Niveles de colesterol	182	207	219		
Índice de masa corporal	25	30	33		

APLICACIONES OPCIONALES DE COMPUTADORA PARA EL CAPÍTULO 4

En el sitio web www.mhhe.com/ritchey2, en *The Statistical Imagination*, están disponibles ejercicios computarizados opcionales del capítulo. Estos ejercicios incluyen la generación de estadísticos de tendencia central con *SPSS Windows* y el uso de la salida para incrementar el sentido de proporción acerca de las formas de distribuciones de puntuaciones para variables de intervalo/razón. Las estadísticas de tendencia central se pueden calcular usando el comando *Descriptives* o el comando *Frequencies*, que se describen en el apéndice D.

Medición de la dispersión o variación en una distribución de puntuaciones

RESUMEN DEL CAPÍTULO

Introducción 136	¿Por qué se llama desviación "estándar"? 148
El rango 138	Puntuaciones estandarizadas (puntuaciones Z) 148
Limitaciones del rango: situaciones en las que reportarlo solo puede conducir a errores 139	La desviación estándar y la distribución normal 150
La desviación estándar 139	Presentación tabular de resultados 153
Pensamiento proporcional y lineal sobre la desviación estándar 140	Insensatez y falacias estadísticas: ¿qué indica cuando la desviación estándar es más grande que la media? 154
Limitaciones de la desviación estándar 145	
La desviación estándar como parte integral de la estadística inferencial 147	

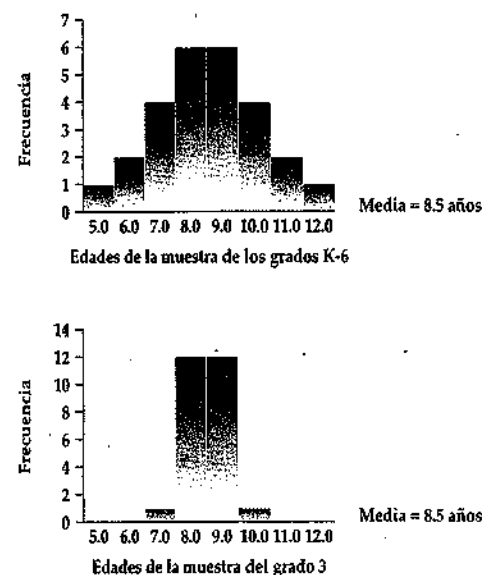
Introducción

Para una variable de intervalo/razón, reportar un estadístico central por sí mismo no es suficiente para comunicar la forma de una distribución de puntuaciones. Dos muestras con las mismas medias pueden tener formas sumamente diferentes. La figura 5-1 presenta dos distribuciones de edades: para una muestra de alumnos de escuela primaria (desde jardín de niños hasta sexto grado, o K-6) y un grupo de tercer grado de otra escuela. La edad media de los alumnos de ambas escuelas es de 8.5 años. En la escuela K-6, sin embargo, los niños tienen entre 5 y 12 años; en el tercer grado de la otra escuela ninguno de los alumnos es menor de 7 años ni mayor de 10 años. Aun cuando estas dos distribuciones de edades tienen la misma tendencia central, sus puntuaciones se dispersan de manera muy diferente, con una mayor dispersión de edades en la escuela K-6.

El tema de este capítulo es la **dispersión**, es decir, *cómo se dispersan las puntuaciones de una variable de intervalo/razón de menor a mayor y la forma de la distribución de éstas*. Existe un número infinito de posibles formas de distribución para una variable con una media dada. Todas las puntuaciones podrían agruparse alrededor de la media con la clara forma de una curva de campana, aunque la curva podría ser de diferentes tamaños según el tamaño de la muestra, o bien, las puntuaciones podrían estar ligeramente o muy sesgadas hacia un lado. Además, una sola variable puede tener dispersiones muy diferentes de

FIGURA 5-1

Comparación de la dispersión de las edades de los alumnos de dos muestras con las mismas medias



una población a otra. Por ejemplo, el ingreso familiar anual de residentes en Estados Unidos varía desde cero hasta decenas de millones de dólares, mientras que el ingreso familiar de los pobres que viven en proyectos habitacionales va de cero a unos pocos miles de dólares.

Dispersión Forma en que se dispersan las puntuaciones de una variable de intervalo/razón de menor a mayor y la forma de la distribución entre éstas.

Los **estadísticos de dispersión** describen cómo se dispersan las puntuaciones de una variable de intervalo/razón a lo largo de su distribución. Los estadísticos de dispersión permiten descripciones precisas de la frecuencia de casos en cualquier punto de una distribución. Por ejemplo, si el gobierno federal decide aumentar los impuestos para los "ricos", empleando estadísticos de dispersión podemos identificar el nivel de ingresos del 5 por ciento más ricos de todas las familias del país. Del mismo modo, si un programa de asistencia social se planea para cubrir sólo 10 000 familias de la ciudad, podemos establecer qué nivel de ingreso familiar satisface los requisitos para recibir la asistencia. Estudiar la dispersión es como ir y venir en un paseo por el eje X de un histograma y observar dónde se concentran los casos. ¿La mayor parte de los casos caen alrededor de la media o están cargados hacia algún lado? ¿Cuántos casos caen entre dos puntos? ¿Qué valor de la variable se lleva el 10 por ciento de los casos? Los dos estadísticos de dispersión que más se emplean son el rango y la desviación estándar.

Estadísticos de dispersión Son estadísticos que describen cómo se dispersan las puntuaciones de una variable de intervalo/razón a lo largo de su distribución.

El rango

El **rango** es una expresión de cómo las puntuaciones de una variable de intervalo/razón se distribuyen de menor a mayor, es decir, es la distancia entre las puntuaciones mínima y máxima de una muestra. Se calcula como la diferencia entre las puntuaciones máxima y mínima, más el valor de la unidad de redondeo. El valor de la unidad de redondeo (1, por ejemplo, si las puntuaciones se redondean al número entero más cercano, 0.1 si las puntuaciones se redondean al décimo más cercano, y así sucesivamente) se suma para considerar el límite real inferior de la puntuación más baja y el límite real superior de la puntuación más alta.

Cálculo del rango de una variable X de intervalo/razón

1. Ordena las puntuaciones de la distribución de menor a mayor.
2. Identifica las puntuaciones mínima y máxima.
3. Identifica el valor de la unidad de redondeo (véase el apéndice A como repaso).
4. Calcula el rango:

Rango = (puntuación máxima – puntuación mínima) + valor de la unidad de redondeo

El rango Es una expresión de la forma en que las puntuaciones de una variable de intervalo/razón se distribuyen de menor a mayor.

Calculemos el rango en un ejemplo. Supongamos que X = edad (redondeada al año más cercano) y tenemos la siguiente distribución de puntuaciones:

21, 23, 43, 26, 20, 21, 25

Empieza por ordenar las puntuaciones:

20, 21, 21, 23, 25, 26, 43

Identifica las puntuaciones mínima y máxima de 20 y 43, respectivamente, y distingue que la unidad de redondeo es 1.

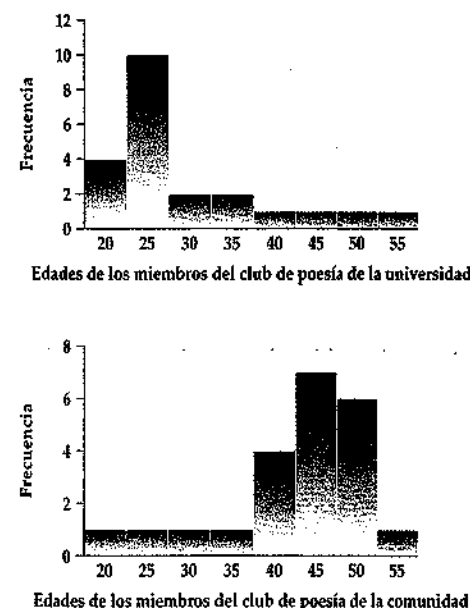
Calcula el rango:

Rango = (puntuación máxima – puntuación mínima) + valor de la unidad de redondeo
 $= (43 - 20) + 1 = 24$ años

Como resultado del redondeo, el individuo que reportó 20 podría tener 19.5 años; y el de 43 años podría tener 43.5 años. El rango de 24 años es la distancia entre estos límites reales menor y mayor de las puntuaciones, es decir, 43.5 años – 19.5 años = 24 años.

FIGURA 5-2

Comparación de dos distribuciones con formas diferentes que tienen el mismo rango



A veces resulta más informativo reportar las puntuaciones mínima y máxima por sí mismas, señalando que estas edades varían desde 20 hasta 43. De esta manera, indirectamente indicamos que en la muestra no hay menores de 20 años ni mayores de 43 años de edad.

Limitaciones del rango: situaciones en las que reportarlo solo puede conducir a errores

Puesto que el rango utiliza las puntuaciones más extremas de una distribución, un valor aislado inflará enormemente su cálculo. Esto sucedió para las siete edades indicadas anteriormente. Los 43 años hicieron que el rango pareciera estar extendido por encima de los 24 años. Reportar esto daría la impresión de que la muestra tiene un número considerable de sujetos de 30 y 40 años. Un reporte más exacto estipularía que, con excepción del estudiante de 43 años, las edades tenían un rango de 7 años ($26 - 20 + 1 = 7$ años). Omitir el valor aislado e indicarlo como excepción es una forma razonable de ajustar esta limitación del rango.

El rango también está limitado por su estrecho alcance informativo. No nos dice nada sobre la forma de la distribución entre las puntuaciones extremas. Por ejemplo, las dos distribuciones descritas en la figura 5-2 tienen el mismo rango, lo que sugiere formas similares, pero de hecho sus formas son radicalmente diferentes. Por último, hay poco que pueda hacerse matemáticamente con el rango. En suma, el rango tiene utilidad limitada, en especial cuando se reporta solo.

La desviación estándar

La desviación estándar es otra medida sumaria de la dispersión o variación de las puntuaciones de una distribución. Este estadístico de dispersión es muy diferente del rango. Al concentrarse en los extremos de la distribución, el rango se aproxima a la dispersión desde

“fuera” o desde los extremos de la distribución. Observar el rango es como ver un juego de baloncesto desde lo alto de las tribunas; la cancha parece encajonada por los tableros de cada extremo. En contraste, la **desviación estándar** describe la forma en que las puntuaciones de una variable de intervalo/razón se dispersan a lo largo de la distribución en relación con la puntuación media. La media es un estadístico de tendencia central y como tal proporciona un punto de enfoque que se centra “dentro” de la distribución. Observar la dispersión a partir de la media con su desviación estándar es como mirar desde el centro de la cancha; el centro de atención está en la distancia del centro de la cancha a otros puntos en cualquier dirección. Al igual que la media, la desviación estándar es muy apropiada con variables de intervalo/razón.

La desviación estándar Describe la forma en que las puntuaciones de una variable de intervalo/razón se dispersan por la distribución en relación con la puntuación media.

Pensamiento proporcional y lineal sobre la desviación estándar

Para una variable de intervalo/razón, la desviación estándar se calcula determinando qué tan alejada está cada puntuación de la media, es decir, cuánto se *desvía* de la media. En este sentido, la desviación estándar es una derivada (o producto) de la media, y las dos medidas siempre se reportan juntas. De hecho, la frase “la media y la desviación estándar” es una de las más empleadas por los estadísticos. La desviación estándar, como una medida sumaria de todas las puntuaciones de una distribución, nos dice con qué amplitud se agrupan las puntuaciones alrededor de la media. Como brevemente lo analizaremos, la desviación estándar también es útil en conjunción con la curva normal.

La siguiente es la fórmula para calcular la desviación estándar:

Cálculo de la desviación estándar

$$s_X = \sqrt{\frac{\sum (X - \bar{X})^2}{n - 1}}$$

donde

s_X = desviación estándar para la variable X de intervalo/razón

$\bar{X}$ = media de X

n = tamaño muestral

Merece la pena seguir un método de paso a paso al cálculo de la desviación estándar. Esto elimina el misterio de la fórmula (con sus símbolos de Σ , cuadrado y raíz cuadrada) y nos ayuda a apreciar que la desviación estándar es parte esencial de la curva normal.

Identifica especificaciones Comenzamos por identificar la información dada.

Especificación: X = una variable de intervalo/razón, n = tamaño muestral, y una distribución de puntuaciones en bruto para X .

Calcula la media Calculamos la media porque la desviación estándar está diseñada para medir la dispersión alrededor de la media.

$$\bar{X} = \frac{\sum X}{n}$$

Calcula las puntuaciones de desviación: pensamiento lineal A continuación determinamos qué tan alejada está la puntuación de cada individuo respecto a la media. La diferencia entre una puntuación y su media se llama **puntuación de desviación**, es decir, cuánto difiere o se “desvía” de la media una puntuación individual:

$$X - \bar{X} = \text{puntuación de desviación para un valor de } X$$

Considere una puntuación de desviación como una medida de distancia en el eje X . ¿Qué nos dice la puntuación de desviación? Supongamos que X es la variable *peso*, y el peso medio es de 138 libras para una muestra de jugadoras de voleibol de la Universidad de Elmstown. La jugadora estrella, Sandra “Mi Amor” Carson, pesa 173 libras; ésta es su puntuación en bruto o “puntuación X ”. Su puntuación de desviación es 35 libras:

$$\text{Puntuación de desviación} = X - \bar{X} = 173 - 138 = 35 \text{ libras}$$

La puntuación de desviación nos dice dos cosas sobre una puntuación de la distribución: (1) la cantidad o distancia a la que la puntuación X se aleja de la media y (2) la dirección de la puntuación X : si está abajo o arriba de la media. Cuando una puntuación X es mayor que la media, la puntuación de desviación resultará un valor positivo, como el de Sandra, lo cual significa que la puntuación X se encuentra a la derecha de una curva de distribución. Cuando una puntuación X es menor que la media, la puntuación de desviación resultará negativa, lo que significa que la puntuación X queda a la izquierda de la media. La puntuación de desviación de Sandra de +35 libras nos indica que ella está 35 libras *por encima* del peso medio del equipo.

Puntuación de desviación Indica cuánto es que una puntuación individual difiere o “se desvía” de la media.

La puntuación de desviación es el cálculo matemático central para determinar la desviación estándar. Como una medida breve para toda la muestra, la desviación estándar es una suma y promedio del cuadrado de estas puntuaciones de desviación, como en los pasos siguientes.

Suma las puntuaciones de desviación El siguiente paso para calcular la desviación estándar es sumar las puntuaciones de desviación. Esta suma siempre será igual a cero (dentro del error de redondeo):

$$\sum (X - \bar{X}) = 0 = \text{suma de las puntuaciones de desviación}$$

La suma de las puntuaciones de desviación es una verificación respecto a la exactitud de los cálculos, porque la suma de las puntuaciones de desviación *siempre* será igual a cero (den-

tro del error de redondeo). En el capítulo 4 vimos la forma en que la media es un punto de equilibrio en la distribución. Lo que hace la media es balancear las desviaciones, para que se cancelen entre sí y resulten en una suma de puntuaciones de desviación igual a cero. De hecho, otra definición matemática de la media es *aquel punto en una distribución donde las puntuaciones de desviación suman cero*.

Eleva al cuadrado las puntuaciones de desviación y suma los cuadrados La dispersión de una variable a menudo se compara para dos o más muestras. El hecho de sumar las puntuaciones de desviación no detectará una diferencia en la dispersión entre dos muestras, porque la suma para ambas será cero. Esto potencialmente nos deja en un callejón sin salida. Si las puntuaciones de una muestra se dispersan ampliamente y si las de la otra lo hacen de manera estrecha, ¿qué beneficio implica informar que ambas tienen una suma de puntuaciones de desviación de cero? ¡Ninguno! Por consiguiente, al comparar dos muestras, debemos encontrar una manera de sumar las puntuaciones de desviación para que la suma sea más grande para una muestra con una dispersión mayor. La solución más útil consiste en elevar al cuadrado cada puntuación de desviación y después sumar los cuadrados. Al elevar al cuadrado se eliminan los signos negativos en las puntuaciones de desviación. La *suma de las puntuaciones de desviación al cuadrado es la variación* (a menudo se denomina *suma de cuadrados*), un estadístico que resume las desviaciones para toda la muestra:

$$\Sigma (X - \bar{X})^2 = \text{la variación (o "suma de cuadrados")}$$

La variación o suma de cuadrados Es la suma de las puntuaciones de desviación al cuadrado; un estadístico que resume las desviaciones para la muestra entera.

Divide la suma de cuadrados entre $n - 1$ para ajustar el tamaño y el error de la muestra: pensamiento proporcional La suma de cuadrados, o variación, constituye una buena medida de la dispersión de una distribución, pero este estadístico presenta dos problemas. Primero, supongamos que deseamos comparar las distribuciones de dos muestras de tamaños diferentes. Por ejemplo, podemos comparar las distribuciones de los promedios para muestras de estudiantes de la Crosstown University ($n = 88$) y de la universidad local ($n = 104$). Cuando sumamos los cuadrados para cada muestra, podría suceder que obtuviéramos una suma más alta para la universidad estatal, simplemente porque sumamos más números, 104 casos en lugar de sólo 88. Cada puntuación X añade cierta cantidad al cálculo. En otras palabras, todo lo demás es igual, y cuanto más observaciones existan, mayor será la suma de cuadrados. Para realizar una comparación equilibrada de dos muestras de tamaño diferente, entonces, necesitamos ajustar el número de observaciones en cada muestra dividiendo cada uno entre su tamaño muestra (n). Esto nos da la *variación promedio* (la media de la suma de cuadrados) en cada muestra. De esta forma ajustamos la suma de cuadrados en proporción al número de casos en la muestra.

Una segunda consideración respecto al tamaño de la muestra es que incluirá el error de muestreo; cuanto mayor sea la muestra, menor será el error de muestreo. Los estadistas han determinado que si restamos 1 de n , este pequeño ajuste produce un estadístico de la muestra que estima con mayor precisión el parámetro de la población. Dicho en otras palabras, si restamos 1 del tamaño de la muestra, realizamos un ajuste para el error de muestreo. (Considera

que con muestras grandes este ajuste tendría poco efecto en el cálculo, mientras que con muestras pequeñas tendría un gran efecto.)

En resumen, dividimos la variación (suma de cuadrados) entre $n - 1$ para compensar tanto los efectos del tamaño muestral de la suma como el error de muestreo. El resultado se llama *varianza*, y su símbolo es s_x^2 :

$$s_x^2 = \frac{\Sigma (X - \bar{X})^2}{n - 1} = \text{varianza de una muestra}$$

La **varianza** es la *variación promedio de las puntuaciones en una distribución*. Para evitar confundir la varianza y la variación nota el sonido acentuado en "varianza" y advierte que n está en su denominador. (Finalmente, debemos hacer notar que si la desviación estándar se calcula para las puntuaciones de una población entera el error de muestreo no constituirá un problema. Por consiguiente, no necesitamos restar 1 de n para obtener la variación de una población, que se simbolizaría como σ_x^2 .)

La varianza Es la variación promedio de las puntuaciones en una distribución (es decir, la media de la suma de cuadrados).

Saca la raíz cuadrada de la varianza para obtener la desviación estándar Para producir una buena medida de dispersión se requiere un último paso. La varianza es perfectamente aceptable para cálculos, pero no se interpreta de manera directa porque las unidades de medida están elevadas al cuadrado. Así, podríamos calcular la varianza de peso para el equipo de fútbol de la universidad local, y encontraríamos que es de 1 391.45 libras cuadradas. Bien, ¿qué es una "libra cuadrada"? Es una libra multiplicada por una libra, pero excepto quizás un matemático ¿quién sabe lo que realmente significa? Necesitamos una unidad de medida directamente interpretable, libras en lugar de libras al cuadrado. Para "regresar" a libras, sacamos la raíz cuadrada de la varianza. (La raíz cuadrada de una unidad de medida al cuadrado es la unidad de medida en sí.) El resultado es la *desviación estándar*:

$$s_x = \sqrt{\frac{\Sigma (X - \bar{X})^2}{n - 1}} = \sqrt{s_x^2}$$

En el caso del peso del equipo local la desviación estándar sería 37.30 libras:

$$\begin{aligned} s_x &= \sqrt{\frac{\Sigma (X - \bar{X})^2}{n - 1}} = \sqrt{s_x^2} \\ &= \sqrt{1\,391.45} = 37.30 \text{ libras} \end{aligned}$$

Los elementos de la ecuación de la desviación estándar —es decir, las puntuaciones de desviación, la suma de los cuadrados o variación y la varianza— son importantes por sí mismos. Estos elementos aparecen por sí mismos en numerosas fórmulas estadísticas (véase,

por ejemplo, el capítulo 12). Los pasos para calcular la desviación estándar se resumen en la tabla 5-1, que el lector encontrará sumamente útil en capítulos posteriores.

Es buena práctica elaborar una hoja de trabajo para estos cálculos. La tabla 5-2 presenta una hoja de trabajo para calcular la desviación estándar de los pesos de 12 de los 98 jugadores del equipo de fútbol de Crosstown.

Para calcular las puntuaciones de desviación, $X - \bar{X}$, calculamos la media y restamos cada puntuación de ella para obtener la tercera columna de la hoja de cálculo:

$$\bar{X} = \frac{\sum X}{n} = \frac{2\,856}{12} = 238 \text{ libras}$$

TABLA 5-1 Para comprender la desviación estándar por medio de su cálculo

Pasos en el cálculo de la desviación estándar	Lo que obtiene el paso
1. Identifica las especificaciones.	1. X debe ser una variable de intervalo/razón.
2. Calcula la media: $\bar{X} = \frac{\sum X}{n}$	2. Porque la desviación estándar está basada en las desviaciones desde la media.
3. Calcula las puntuaciones de desviación: $X - \bar{X}$	3. Para determinar la distancia de cada puntuación desde la media.
4. Suma las puntuaciones de desviación: $\sum (X - \bar{X})$	4. Asegúrate de que $\sum (X - \bar{X}) = 0$
5. Eleva al cuadrado las puntuaciones de desviación y súmalas para obtener la variación de la suma de cuadrados: $\text{Variación} = \sum (X - \bar{X})^2$	5. Las puntuaciones de desviación se elevan al cuadrado para eliminar signos negativos y obtener una suma diferente de cero.
6. Calcula la varianza: $s_x^2 = \frac{\sum (X - \bar{X})^2}{n - 1}$	6. Divide entre $n - 1$ la suma de cuadrados para ajustar el tamaño muestral y el error de redondeo.
7. Calcula la desviación estándar, s_x : $s_x = \sqrt{\frac{\sum (X - \bar{X})^2}{n - 1}} = \sqrt{s_x^2}$	7. Toma la raíz cuadrada de la varianza para obtener directamente unidades de medida que se puedan interpretar (unidades en lugar de unidades al cuadrado).

TABLA 5-2 Hoja de trabajo para calcular la desviación estándar: peso de jugadores de fútbol de Crosstown ($n = 12$)

Especificaciones		Cálculos	
(1) Jugador	(2) X	(3) $X - \bar{X}$	(4) $(X - \bar{X})^2$
1	165	-73	5 329
2	200	-38	1 444
3	216	-22	484
4	217	-21	441
5	226	-12	144
6	236	-2	4
7	239	1	1
8	244	6	36
9	261	23	529
10	268	30	900
11	283	45	2 025
12	301	63	3 969
$n = 12$	$\sum X = 2\,856$ libras	$\sum (X - \bar{X}) = 0$	$\sum (X - \bar{X})^2 = 15\,306$ libras al cuadrado

Por último, elevamos al cuadrado las puntuaciones de desviación de la columna 3 para obtener la columna 4. La suma de la columna 4 de la tabla 5-2 y el tamaño de la muestra n son todo lo que necesitamos para calcular la desviación estándar:

$$s_x = \sqrt{\frac{\sum (X - \bar{X})^2}{n - 1}} = \sqrt{\frac{15\,306}{11}} = \sqrt{1\,391.45} = 37.30 \text{ libras}$$

Limitaciones de la desviación estándar

Como la desviación estándar se calcula a partir de la media, al igual que ésta se infla por los valores extremos. Éstos generan puntuaciones con grandes desviaciones. Cuando se elevan al cuadrado, estas grandes desviaciones, ya sean positivas o negativas, producen un alto resultado positivo e inflado. Así, la desviación estándar puede ser muy confusa cuando se reporta para una distribución sesgada, en la que pocas puntuaciones se extienden en una dirección. Para convencerse del efecto de las puntuaciones extremas tanto en la media como en la desviación estándar, completa la hoja de cálculo de la tabla 5-2; pero agrega los dos casos siguientes para obtener una nueva muestra con $n = 14$: el jugador 13 que pesa 115 libras y el jugador 14 que pesa 125 libras. A continuación compara las respuestas de las muestras original y nueva.

Para calcular el rango y la desviación estándar

Problema: calcula el rango, media y desviación estándar de los impuestos de gasolina cobrados por 10 estados seleccionados del oeste de Estados Unidos. Estos estados y sus impuestos se presentan en la tabla 5-3, donde X = impuesto de gasolina por galón.

TABLA 5-3 | Impuestos estatales a la gasolina en estados seleccionados del oeste en mayo de 1996.

Especificaciones		Cálculos	
Estado	Impuesto (¢) por galón X	$X - \bar{X}$	$(X - \bar{X})^2$
Nuevo México	17	-4.7	22.09
California	18	-3.7	13.69
Arizona	18	-3.7	13.69
Utah	19	-2.7	7.29
Colorado	22	0.3	0.09
Washington	23	1.3	1.69
Nevada	23	1.3	1.69
Oregón	24	2.3	5.29
Idaho	25	3.3	10.89
Montana	28	6.3	39.69
$\Sigma X = 217¢$		$\Sigma(X - \bar{X})^2 = 116.10¢$ al cuadrado	
$n = 10$		$\Sigma(X - \bar{X}) = 0$	

Fuente: Tarifas de impuestos de <http://www.api.org/news/596stax.htm>. Copyright © 1996 por American Petroleum Institute. Reimpreso con permiso del Instituto.

1. Asegúrate de que la variable sea de nivel de medición de intervalo/razón (como en el caso de los impuestos a la gasolina).
2. Organiza los datos en una hoja de cálculo con puntuaciones ordenadas de menor a mayor y con los siguientes encabezados de columna:

Caso	X	$X - \bar{X}$	$(X - \bar{X})^2$
------	-----	---------------	-------------------

donde "Caso" = número o nombre del caso,

X = puntuación observada, sin elaborar, de la variable de intervalo/razón,

$X - \bar{X}$ = puntuación de desviación,

$(X - \bar{X})^2$ = cuadrado de la puntuación de desviación.

3. Calcula el rango. Con estas puntuaciones ordenadas, vemos que la puntuación mínima es 17 centavos y la máxima es 28 centavos. Nuestra unidad de redondeo es un número entero.

Rango = (puntuación máxima - puntuación mínima) + valor de unidad de redondeo

$$= (28 - 17) + 1 = 12¢$$

4. Calcula la media.

$$\bar{X} = \frac{\Sigma X}{n} = \frac{217}{10} = 21.7¢$$

5. A partir de la media, resta cada una de las puntuaciones en bruto (X) para obtener su puntuación de desviación. Suma las puntuaciones de desviación para asegurarte de que totalicen cero (dentro del error de redondeo).
6. Resta cada una de las puntuaciones de desviación y suma los cuadrados.
7. Calcula la desviación estándar:

$$s_x = \sqrt{\frac{\Sigma(X - \bar{X})^2}{n - 1}} = \sqrt{\frac{116.10}{9}} = 3.59¢$$

La desviación estándar como parte integral de la estadística inferencial

Las características de la media y de la desviación estándar las hacen muy útiles para alcanzar un sentido de proporción respecto de las variables individuales que se estudian. La desviación estándar y las puntuaciones de desviación, a partir de las cuales se calculan, también son esenciales para examinar las relaciones entre dos variables. El foco de la estadística inferencial consiste en desarrollar una comprensión de por qué las puntuaciones individuales de una variable dependiente se desvían de su media.

Supongamos, por ejemplo, que estamos estudiando el abuso en el consumo de alcohol. Para una muestra de bebedores adultos, encontramos que la media del consumo de bebidas alcohólicas es de 4.3 galones por año. Gary consumió 7.3 galones el último año, 3 galones arriba de la media. Sam consumió sólo 1 galón, 3.3 galones abajo de la media. ¿Qué sucede con estas desviaciones alta y baja? Quizá podríamos generar hipótesis acerca de algunas variables predictoras (independientes) que creamos que estén relacionadas con esta variable dependiente. Por ejemplo, la hipótesis del consumo a la hora de la comida podría explicar, en parte, la puntuación de desviación positiva de Gary, los bebedores de familias que consumen vino con sus alimentos tienen un consumo de alcohol medio más alto. Existe también la hipótesis del bebedor social, la cual podría explicar, en parte, la puntuación de desviación negativa de Sam, los bebedores que sólo consumen alcohol en reuniones sociales tienen un consumo de alcohol medio más bajo.

Para una muestra completa, nuestro interés radica en explicar la variación, la suma de puntuaciones de desviación al cuadrado. Las puntuaciones de desviación, la variación y la desviación estándar simplemente son medidas de diferencias en las puntuaciones para una

variable entre los sujetos de una población. ¿Es más alta la cantidad media de consumo de alcohol anual para las personas de ciertas regiones, entre diferentes edades o grupos religiosos o entre sexos? Las respuestas a tales preguntas dependen de las propiedades matemáticas de la media, la desviación estándar y la curva normal.

¿Por qué se llama desviación “estándar”?

La desviación estándar recibe su nombre por el hecho de que proporciona una *unidad de medida común* (un estándar) para comparar variables con *unidades observadas de medida* muy diferentes. Por ejemplo, imagine que Mary Smith y Jason Jones solicitan una beca con base en su desempeño en los exámenes de admisión a la universidad. Mary contestó la prueba académica de la universidad (ACT) y obtuvo 26 puntos ACT. Jason hizo lo propio con la prueba de admisión Stanford (SAT) y obtuvo 900 puntos SAT. Estos dos resultados de las pruebas tienen unidades de medida muy diferentes: los puntos de la prueba ACT van de cero a 36; y los de la prueba SAT, de 200 a 1 600. Las puntuaciones en bruto para las dos pruebas no pueden compararse directamente, pero con el uso de las medidas y las desviaciones estándar para ambas pruebas podemos crear una manera para compararlas. Con los siguientes estadísticos, encontramos que, en comparación con otros aspirantes que contestan las pruebas, Mary obtuvo la puntuación más alta:

$$\begin{array}{lll} X = \text{puntuación de la prueba ACT} & \bar{X} = 22 \text{ puntos ACT} & s_X = 2 \text{ puntos ACT} \\ Y = \text{puntuación de la prueba SAT} & \bar{Y} = 1000 \text{ puntos SAT} & s_Y = 100 \text{ puntos SAT} \end{array}$$

La puntuación de ACT de 26 que obtuvo Mary tiene una desviación estándar de 2 arriba de la media de aquellos que toman la prueba ACT, es decir, su puntuación está 4 puntos ACT, esto es, 2 por 2 desviaciones estándar sobre el promedio de 22. La puntuación de Jason es de 1 desviación estándar abajo de la media de aquellos que contrastan la prueba SAT, es decir, su puntuación está 100 puntos SAT, 1 desviación estándar abajo del promedio de 1 000. Sin lugar a dudas podemos otorgarle la beca a Mary. Utilizando las desviaciones estándar como unidades de medida en lugar de “puntos de prueba ACT” y de “puntos de prueba SAT”, tenemos una norma común o estándar para ambas variables, de ahí el nombre de *desviación estándar*. ¿Quién te dijo que no podías comparar peras con manzanas?

Puntuaciones estandarizadas (puntuaciones Z)

El ejemplo anterior ilustra el hecho de que la puntuación de un sujeto de la investigación en cualquier variable de intervalo/razón puede expresarse de diversas maneras. Primero, lo expresamos en sus unidades de medida observadas, originales, como una **puntuación en bruto**. Por ejemplo, la puntuación en bruto X de Mary es 26 puntos ACT. Segundo, lo expresamos como una **desviación de la media**, es decir, la puntuación de desviación $(X - \bar{X})$; la puntuación de desviación de Mary es +4 y significa que ella obtuvo 4 puntos ACT arriba de la media de aquellos que tomaron el ACT. Tercero, expresamos su puntuación como un **número de desviaciones estándar de la media** de la puntuación ACT. Llamamos a esto su **puntuación estandarizada** (o **puntuación Z**), que para la variable X se calcula como sigue:

Cálculo de puntuaciones estandarizadas (puntuaciones Z)

$$Z_X = \frac{X - \bar{X}}{s_X}$$

donde

- Z_X = puntuación estandarizada para un valor de X
- = número de desviaciones estándar que una puntuación en bruto (puntuación X) se desvía de la media
- X = una variable de intervalo/razón
- $\bar{X}$ = la media de X
- s_X = la desviación estándar de X

Si hacemos que la puntuación $X = \text{ACT}$ con $\bar{X} = 22$ puntos ACT y $s_X = 2$ puntos ACT, la puntuación Z de Mary es

$$Z_X = \frac{X - \bar{X}}{s_X} = \frac{26 - 22}{2} = \frac{4}{2} = 2.00 \text{ SD}$$

donde SD significa “desviaciones estándar”. Una puntuación Z es la distancia de una puntuación X hacia la media (es decir, su puntuación de desviación) dividida entre la desviación estándar de las distancias.

Una clave para tener claras estas tres maneras de expresar la puntuación consiste en enfocarse en las unidades de medida. Las puntuaciones en bruto y las puntuaciones de desviación para una variable se presentan en la unidad de medida original observada, que, por supuesto, es definida por una variable. Por ejemplo, la unidad de medida observada para edad es años; para peso, libras o kilogramos; para altura, pulgadas o centímetros; y así sucesivamente. Pero cualquiera que sea la unidad de medida de una variable, sus puntuaciones Z se miden en SD. La tabla 5-4 resume estas distinciones.

Aquí aparecen algunos ejemplos de una muestra aleatoria de mujeres estudiantes en la universidad local:

1. Donde X = peso, $\bar{X} = 120$ libras, $s_X = 10$ libras:

Caso	X (peso)	$X - \bar{X}$ (puntuación de desviación)	Z_X (puntuación estandarizada)
Cheryl Jones	110 libras	-10 libras	-1 SD
Jennifer Smith	125 libras	5 libras	.5 SD
Terri Barnett	107 libras	-13 libras	-1.3 SD

2. Donde Y = estatura, $\bar{Y} = 65$ pulgadas, $s_Y = 3$ pulgadas:

Caso	Y (estatura)	$Y - \bar{Y}$ (puntuación de desviación)	Z_Y (puntuación estandarizada)
Cheryl Jones	64 pulgadas	-1 pulgada	-.33 SD
Jennifer Smith	65 pulgadas	0 pulgada	0 SD
Terri Barnett	68 pulgadas	3 pulgadas	1 SD

TABLA 5-4 | Diferentes formas en las que pueden presentarse puntuaciones de una variable

Forma de puntuación para una variable y su símbolo	Unidades de medida de la variable	Ejemplo: X = estatura
Puntuación en bruto (puntuación X): X	Unidad de medida de la variable	Pulgadas
Puntuación de desviación = $X - \bar{X}$	Unidad de medida de la variable	Pulgadas
Puntuación estandarizada (Z_x) o "puntuación Z ": $Z_x = \frac{X - \bar{X}}{s_x}$	Desviaciones estándar de la variable (SD)	SD

Recuerda que las puntuaciones de desviación y las puntuaciones Z son medidas de la distancia desde la puntuación en bruto de la variable hasta su media. La puntuación de desviación se obtiene restando la media de la puntuación en bruto (es decir, $X - \bar{X}$). Al dividir esta puntuación de desviación entre la desviación estándar, cortamos esta puntuación de desviación en las partes y múltiplos de las desviaciones estándar desde la media. Recuerda que después de calcular la media, calcular las puntuaciones de desviación es lo siguiente que hacemos cuando calculamos la desviación estándar. La esencia de la desviación estándar está en ver una puntuación en bruto individual como una desviación desde la media.

Para obtener un buen sentido de proporción sobre las fórmulas para las puntuaciones de desviación y las puntuaciones Z , examinemos las relaciones entre los tamaños de las puntuaciones en bruto, las puntuaciones de desviación y las puntuaciones Z . Primero, cuanto más lejana de la media esté una puntuación X mayores serán su puntuación de desviación y su puntuación Z . Primero, cuanto más lejana de la media esté una puntuación X mayores serán su puntuación de desviación y su puntuación Z . Es más, el signo de cualquier puntuación de desviación y puntuación Z indica la **dirección de una puntuación**: ya sea que la observación caiga arriba de la media (la dirección positiva) o debajo de la media (la dirección negativa). El signo “-” (signo menos) indica que una puntuación en bruto está debajo de la media; el signo “+” (signo más), que está implícito, no escrito, indica que está encima de la media. En los ejemplos anteriores Cheryl y Terri están abajo del promedio en peso, y Terri está arriba del promedio en estatura. De hecho, a partir de estas puntuaciones Z podemos decir que Terri es una persona alta, delgada más de 1 SD abajo en peso, pero 1 SD arriba en estatura. Jennifer tiene estatura media; así, su puntuación de desviación y su puntuación Z para Y son cero: ella no se desvía de la estatura media.

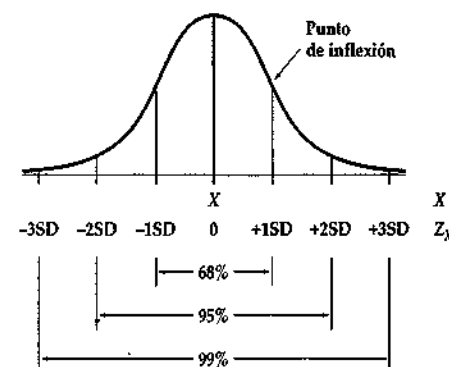
Puesto que usaremos puntuaciones Z o medidas similares de desviación en cada capítulo en el resto del texto, es prudente practicar cómo calcular puntuaciones de desviación y puntuaciones Z , así como estudiar las direcciones (signos) de esas puntuaciones. Se recomienda una doble verificación. Si una puntuación en bruto queda debajo de la media, su desviación y sus puntuaciones Z son negativas. También debes tener presente que las puntuaciones Z son simplemente otra manera de expresar puntuaciones en bruto. Cada puntuación en bruto tiene una puntuación Z correspondiente, y viceversa.

La desviación estándar y la distribución normal

Además de proporcionar un estándar de comparación entre variables y muestras diferentes, bajo condiciones apropiadas la media y la desviación estándar ofrecen gran riqueza de información. Éste es el caso cuando una variable tiene una distribución de puntuaciones que es normal, formada como la curva de distribución normal. Como lo definimos en el capítulo 4, una distribución normal es simétrica, con su media, mediana y moda iguales entre sí y

FIGURA 5-3

Relación entre la desviación estándar y la curva normal



localizadas en el centro de la curva. Sin embargo, la simetría o equilibrio en la curva no es toda la imagen. La curva normal también tiene una forma de campana inconfundible, que no es muy plana ni demasiado puntiaguda. Muchas variables se distribuyen normalmente (por ejemplo, la estatura, peso e inteligencia). Cualquiera que sea la variable que se examine, si está normalmente distribuida tendrá las propiedades de una curva normal.

Lo que hace que una desviación estándar sea una herramienta estadística tan valiosa es que es una parte matemática de la curva normal. Cuando se sigue la curva desde su centro (es decir, su pico) en cualquier dirección, la curva cambia de forma para aproximarse al eje X . Desde el pico, el punto en el que la curva empieza a desplazarse hacia fuera es 1 desviación estándar desde la media. Este punto recibe el nombre de **punto de inflexión** y se destaca en la figura 5-3. Esto indica que la media y la desviación estándar son aspectos matemáticos de un fenómeno natural: la tendencia a que una distribución normal en forma de campana se presente para numerosos eventos naturales.

Comprender el fenómeno de normalidad es un aspecto importante de la imaginación estadística. Muchos fenómenos que ocurren naturalmente tienen distribuciones de frecuencias en forma de campana como la curva normal. La curva normal ilustra el hecho de que cuando nos desviamos más allá de la media esperamos encontrar cada vez menos casos. Para muchas variables, existe un promedio alrededor del cual cae la mayoría de las puntuaciones, y cuando nos alejamos de este promedio, las frecuencias del caso disminuyen. Por ejemplo, la estatura física se distribuye normalmente; la mayoría de las personas están cerca del promedio, con unas cuantas personas muy altas o muy bajas.

Uno de los rasgos más sobresalientes del fenómeno de normalidad, que ocurre naturalmente, es que ofrece predicciones precisas sobre cuántas puntuaciones de una población caen dentro de cualquier rango de puntuaciones. Como se ilustró en la figura 5-3, *para cualquier variable normalmente distribuida*:

1. Cincuenta por ciento de las puntuaciones caen encima de la media; 50 por ciento, debajo. Esto se debe al hecho de que la mediana es igual a la media.
2. Prácticamente todas las puntuaciones caen dentro de 3 desviaciones estándar a partir de la media en ambas direcciones. Ésta es una distancia de 3 puntuaciones Z abajo de la media en ambas direcciones. Ésta es una distancia de 3 puntuaciones Z arriba de la media, una amplitud total de 6 desviaciones estándar. La cantidad precisa es 99.7 por ciento. El restante 0.3 por ciento de casos (es decir, 3 casos de cada 1 000) caen fuera de 3 desviaciones estándar y, teóricamente, la curva se extien-

de hacia el infinito en ambas direcciones. (Prácticamente hablando, las puntuaciones para algunas variables, como el peso corporal, tienen límites finitos.)

3. Cerca del 95 por ciento de las puntuaciones de una variable normalmente distribuida caen dentro de una distancia de 2 desviaciones estándar en ambas direcciones de la media. Esto es más menos 2 puntuaciones Z de la media.
4. Alrededor de 68 por ciento de las puntuaciones de una variable normalmente distribuida caen dentro de una distancia de 1 desviación estándar (más menos 1 puntuación Z) en ambas direcciones de la media.

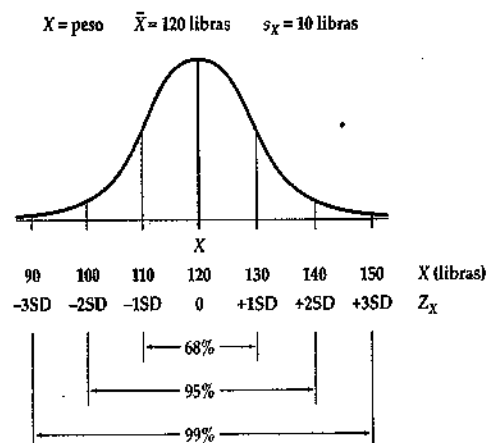
Recuerda que la distribución normal tiene características muy predecibles. Si una variable se distribuye en esta peculiar forma de campana, podemos utilizar los estadísticos de la muestra y lo que sabemos respecto de la curva normal para estimar cuántas puntuaciones en una población caen dentro de cierto rango.

Para ilustrar la utilidad de la curva normal, sigamos nuestro ejemplo: una muestra de mujeres estudiantes de la Universidad Crosstown, donde X = peso, el peso medio es de 120 libras y $s_x = 10$ libras. Primero, necesitamos asegurarnos de que la distribución de las puntuaciones es, de hecho, normal, es decir que tenga forma de campana. Esto podría hacerse elaborando un histograma de las puntuaciones de una muestra (no se ilustra). Si la forma de este gráfico se aproxima a la de una campana, suponemos que esta variable no sólo está normalmente distribuida en la muestra sino también en la población. Nos referimos a este hecho como "suponiendo normalidad". (La forma de un histograma de la muestra puede ser ligeramente fuera de lo normal como resultado del error muestral.) Como se grafica en la figura 5-4, suponiendo normalidad, podemos hacer las siguientes estimaciones de los pesos de la población de mujeres estudiantes de la Universidad Crosstown.

1. La mitad de estas estudiantes pesa más de 120 libras.
2. Cerca del 68 por ciento de las mujeres estudiantes de la Universidad Crosstown pesan entre 110 y 130 libras.

FIGURA 5-4

Uso de la curva normal para estimar la distribución de peso (X) en mujeres estudiantes de la Universidad Crosstown



3. Alrededor del 95 por ciento de las mujeres estudiantes de la universidad local pesan entre 100 y 140 libras.
4. Muy pocas pesan menos de 90 libras o más de 150 libras.

Recuerda, una puntuación Z simplemente es otra forma de expresar una puntuación en bruto (es decir, la puntuación X para una observación individual). Si Susana pesa 110 libras, ella está 1 SD debajo del peso medio y tiene una puntuación Z de -1.00 SD.

Presentación tabular de resultados

En artículos de investigación, una tabla básica de estadística descriptiva es la que lista todas las variables y sus medias y desviaciones estándar. La tabla 5-5 presenta una tabla de estadística descriptiva de un estudio del bienestar psicológico de personas sin hogar en dos puntos en el tiempo.

TABLA 5-5 | Estadísticos descriptivos para síntomas psicológicos, satisfacción con la vida y autoestima.

Subescalas	Tratamiento 1		Tratamiento 2	
	M	SD	M	SD
Síntomas psicológicos				
Enojo	4.17	0.80	4.14	0.85
Ansiedad	3.97	0.79	3.97	0.80
Depresión	3.60	0.76	3.68	0.77
Manía	3.59	0.87	3.68	0.90
Psicosis	4.51	0.72	4.52	0.72
Satisfacción con la vida				
Vestido	4.33	1.59	4.49	1.60
Alimento	4.79	1.53	4.98	1.42
Salud	4.81	1.38	4.77	1.41
Vivienda	4.37	1.49	4.51	1.54
Diversión	3.74	1.53	3.84	1.56
Dinero	2.98	1.57	3.19	1.67
Social	4.42	1.44	4.51	1.79
Autoestima				
Estima 1	3.21	0.85	3.24	0.84
Estima 2	3.36	0.87	3.28	0.85

Nota: $n = 298$. Puntuaciones más altas reflejan mayor bienestar subjetivo.

Fuente: Modificado de Marshall y otros, 1998: 49. Reimpreso con permiso de la American Sociological Association.

Insensatez y falacias estadísticas: ¿qué indica cuando la desviación estándar es más grande que la media?

Como vimos en el capítulo 4, la media es susceptible de distorsión por la presencia de puntuaciones extremas, valores extremos y distribuciones sesgadas. Debido a que se basa en desviaciones desde la media, la desviación estándar es susceptible al mismo problema. La distorsión está determinada por el hecho de que las puntuaciones de desviación están elevadas al cuadrado.

Un tipo común de distribución sesgada es un sesgo positivo (o a la derecha), en el que la mayoría de las personas tienen bajas puntuaciones, pero algunas tienen puntuaciones altas. Por ejemplo, "estancias en el hospital", o el número de veces que una muestra aleatoria de personas de más de 65 años hayan estado en el hospital el año pasado, es un sesgo a la derecha. Casi todas las personas registrarán cero en la estancia; algunas, uno; otras reportarán dos; y pocas personas muy enfermas anotarán estancias frecuentes. Este tipo de distribución se presenta en la tabla 5-6.

TABLA 5-6 Distribución sesgada de estancias en el hospital, durante el último año, entre personas mayores de 65 años (datos ficticios)

Especificaciones		Cálculos	
(1) Caso	(2) X	(3) $X - \bar{X}$	(4) $(X - \bar{X})^2$
1	0	-2.41	5.81
2	0	-2.41	5.81
3	0	-2.41	5.81
4	0	-2.41	5.81
5	0	-2.41	5.81
6	0	-2.41	5.81
7	0	-2.41	5.81
8	0	-2.41	5.81
9	1	-1.41	1.99
10	1	-1.41	1.99
11	1	-1.41	1.99
12	2	-0.41	0.17
13	2	-0.41	0.17
14	5	2.59	6.71
15	9	6.59	43.43
16	10	7.59	57.61
17	10	7.59	57.61

$\Sigma X = 41$ veces

$n = 17$

$\Sigma(X - \bar{X})^2 = 218.15$ veces

$\Sigma(X - \bar{X}) = 0.03^*$

*No totalizó cero debido al error de redondeo.

Incluso sin un histograma, los valores relativos de la media y de la desviación estándar para esta distribución proporcionan una señal de que la distribución está sesgada. Estos estadísticos se calculan como sigue:

X = estancias en el hospital = durante el último año, el número de veces que una persona ingresa en un hospital y pasa ahí por lo menos una noche

$$\bar{X} = 2.41 \text{ veces} \quad s_x = 3.69 \text{ veces} \quad n = 17 \text{ casos}$$

Observa que la desviación estándar es mayor que la media, lo cual sugiere que una o más puntuaciones extremas inflaron la media y la desviación estándar. Además, desde el momento en que se elevan al cuadrado los números de la desviación estándar, unas cuantas puntuaciones extremas pueden hacer "explotar" su valor. Toma nota, por ejemplo, de la enorme contribución a la suma de cuadrados que los tres casos más grandes hicieron con sus estancias de 9, 10 y 10 veces.

¿Por qué una desviación estándar más grande que la media indica un sesgo? Recuerda que si una distribución no está sesgada (es decir, tiene una forma normal de campana), su rango tendrá una amplitud de entre 4 y 6 desviaciones estándar. Cuando la curva se traza, la amplitud de 2 o 3 desviaciones estándar se ajustará en cada lado de la media. Si el límite inferior de las puntuaciones X de una variable es cero, por lo menos la distancia de 2 desviaciones estándar debería ajustarse entre una puntuación X de cero y la media. Cuando la desviación estándar es mayor que la media, como en el caso de las estancias en el hospital, ni una sola amplitud de la desviación estándar puede lograr este ajuste. Otra forma de explicarlo es que la desviación estándar debería ser alrededor de la mitad del tamaño de la media, o menos.

Dos reglas generales se aplican a los tamaños relativos de la media y de la desviación estándar:

1. Si la desviación estándar es más grande que la media, esto probablemente indica un sesgo, es decir, la presencia de valores extremos u otra peculiaridad en la forma de la distribución, por ejemplo una distribución bimodal.
2. Si la desviación estándar no es de la mitad del tamaño de la media o menos, debe tenerse cuidado al examinar la distribución para analizar la posible existencia de sesgos o valores extremos.

Como veremos en capítulos posteriores, cuando una variable sesgada está correlacionada con otras variables, los resultados pueden ser erróneos (capítulo 14). En tales casos, deben realizarse ajustes a los estadísticos para evitar tales errores.

RESUMEN

1. La dispersión se refiere a la forma en que las puntuaciones de una variable de intervalo/razón se dispersan, desde la menor hasta la mayor, y a la forma de la distribución entre ellas. Los estadísticos de dispersión miden esta diseminación.
2. Los estadísticos de dispersión que más se utilizan son el *rango* y la *desviación estándar*.
3. El *rango* es una expresión de la forma en que las puntuaciones de una variable de intervalo/razón se distribuyen de la menor a la mayor. Es la distancia entre las puntuaciones mínima y máxima de una muestra.

4. El rango tiene limitaciones. Es afectado en gran medida por valores extremos. Además tiene un estrecho alcance de información. Indica el ancho de una distribución de puntuaciones, pero no nos dice nada acerca de cómo es que las puntuaciones se dispersan entre las puntuaciones máxima y mínima.
5. La desviación estándar describe el modo en que las puntuaciones de una variable de intervalo/razón se dispersan a lo largo de la distribución, en relación con la puntuación media. La desviación estándar se calcula al determinar qué tan alejada está cada puntuación respecto a la media, es decir, cuánto "se desvía" de la media. Entonces, la desviación estándar está basada en puntuaciones de desviación.
6. La desviación estándar tiene limitaciones. Los valores extremos la inflan en gran medida. Puede ser errónea si la distribución de puntuaciones es sesgada.
7. La desviación estándar indica una unidad estándar de comparación, es decir, una unidad común de medida para comparar variables con unidades de medida muy diferentes. Las puntuaciones estandarizadas (puntuaciones Z) expresan una puntuación en bruto como un número de desviaciones estándar (SD) a partir de la puntuación media. Dos variables con diferentes unidades de medida pueden compararse si ambas están estandarizadas al calcular las puntuaciones Z .
8. La dirección de una puntuación Z está determinada por su signo. Una puntuación Z positiva se presenta cuando una puntuación en bruto es mayor que la media; una puntuación Z negativa ocurre cuando una puntuación en bruto es menor que la media.
9. Hay tres formas de expresar el valor de cualquier puntuación de una variable de intervalo/razón:
 - a) Como una *puntuación en bruto*, el valor observado de X en su unidad de medida original (por ejemplo, pulgadas o libras).
 - b) Como una *puntuación de desviación*, la diferencia entre la media y una puntuación en bruto. Las puntuaciones de desviación también se expresan en la unidad de medida original de la variable.
 - c) Como *puntuación Z* (es decir, la *puntuación estandarizada*), la diferencia entre la media y la puntuación en bruto, pero expresada como un número de desviaciones estándar (SD).
10. La desviación estándar es una parte matemática de la curva normal. Es la distancia en el eje X de la media a la puntuación directamente bajo el punto de inflexión de la curva.
11. Si una variable está normalmente distribuida, podemos usar estadísticos muestrales y lo que sabemos acerca de la curva normal para estimar cuántas puntuaciones de una población caen dentro de cierto rango. a) 50% de las puntuaciones caen arriba de la media y 50% caen abajo. b) Prácticamente todas las puntuaciones (99.7%) caen a no más de tres desviaciones estándar de la media en ambas direcciones. c) Alrededor del 95% de las puntuaciones caen a no más de dos desviaciones estándar de la media en ambas direcciones. d) Alrededor del 68% de las puntuaciones de una variable normalmente distribuida caen a no más de 1 desviación estándar de la media en ambas direcciones.
12. Si la desviación estándar es mayor que la media, la distribución de puntuaciones *no puede* tener forma normal. Es probable que un histograma de la variable deje ver un sesgo o una distribución de puntuaciones de forma extraña.

EXTENSIONES DEL CAPÍTULO EN EL SITIO WEB THE STATISTICAL IMAGINATION

Las extensiones del capítulo 3 del material de texto disponibles en el sitio web *The Statistical Imagination*, en www.mhhe.com/ritchey2, incluyen la forma en que una estimación de la desviación estándar basada en el rango puede usarse para detectar si una distribución de puntuaciones está sesgada.

FÓRMULAS PARA EL CAPÍTULO 5

Organiza una hoja de cálculo con casos en orden:

Especificaciones		Cálculos	
(1) Caso	(2) X	(3) $X - \bar{X}$	(4) $(X - \bar{X})^2$
•	•	...	...
•	•	...	...
•	•	...	...
$\Sigma X = \dots$		$\Sigma(X - \bar{X})^2 = \dots$	
$n = \dots$		$\Sigma(X - \bar{X}) = 0$	

Cálculo del rango:

1. Ordena las puntuaciones de la distribución de menor a mayor.
2. Identifica las puntuaciones mínima y máxima.
3. Identifica el valor de la unidad de redondeo (véase apéndice A).
4. Calcula el rango:

$$\text{Rango} = (\text{puntuación máxima} - \text{puntuación mínima}) + \text{valor de unidad de redondeo}$$

Cálculo de la desviación estándar:

1. Empieza por calcular la media de X y completar una hoja de cálculo semejante a la de la tabla 5-2.
2. Calcula la desviación estándar:

Trabajando con una hoja de cálculo

$$s_X = \sqrt{\frac{\Sigma(X - \bar{X})^2}{n - 1}}$$

Calculando puntuaciones estandarizadas (puntuaciones Z):

$$Z_X = \frac{X - \bar{X}}{s_X}$$

PREGUNTAS PARA EL CAPÍTULO 5

- Los estadísticos de dispersión se calculan ¿sólo en variables de qué niveles de medición?
- Tanto el rango como la desviación estándar son medidas de la dispersión de las puntuaciones en una distribución. Explica las diferencias en perspectiva entre estos dos estadísticos.
- ¿Qué efecto tiene una puntuación extrema o valor aislado sobre el cálculo del rango?
- La desviación estándar se “deriva” de la media. ¿Qué significa esto?
- Al calcular el rango, el valor de la unidad de redondeo de la variable se suma a la diferencia entre las puntuaciones máxima y mínima. ¿Por qué se suma el valor de la unidad de redondeo?
- Al calcular la desviación estándar, ¿por qué es necesario elevar al cuadrado las puntuaciones de desviación?
- Al calcular la desviación estándar para datos de una muestra, ¿por qué debemos dividir entre $n - 1$?
- Al calcular la desviación estándar, ¿por qué se requiere sacar la raíz cuadrada?
- ¿Cuál es la relación matemática entre la varianza y la desviación estándar?
- Menciona otro nombre para la variación.
- ¿Cuál es el significado de la palabra *estándar* en el término *desviación estándar*?
- Una expresión de qué tan lejos está una puntuación en bruto de la media de una distribución, en las unidades de medida originales de la variable X , se llama una puntuación _____.
- Una expresión de qué tan lejos está una puntuación en bruto de la media de una distribución, en unidades de medida de desviaciones estándar (SD), se llama una puntuación _____.
- ¿Cuáles son las propiedades de una distribución normal?
- En una distribución normal, ¿qué porcentaje de puntuaciones caen aproximadamente dentro de 1 desviación estándar de la media en ambas direcciones?, ¿y dentro de 2 desviaciones estándar de la media en ambas direcciones?, ¿y dentro de 3 desviaciones estándar de la media en ambas direcciones?
- En una distribución normal ¿qué porcentaje exacto de puntuaciones caen sobre la media? ¿Qué estadístico de tendencia central, además de la media, explica este fenómeno?
- En una distribución normal la curva alcanza su máximo en el valor de la media. ¿Qué estadístico de tendencia central, además de la media, justifica este fenómeno?
- Si una puntuación en bruto cae debajo de la media en una distribución, el signo de la puntuación Z , ¿será positivo o negativo? Ilustra tu respuesta utilizando la fórmula para calcular una puntuación Z .
- En cualquier distribución de puntuaciones de intervalo/razón hay una puntuación en la cual las desviaciones de ésta suman cero. ¿Qué estadístico de tendencia central se localiza en ese punto?
- Para su grupo de edad Charles está 1 desviación estándar debajo de la estatura media, pero 1.5 desviaciones estándar arriba del peso medio. Describe su complexión corporal general.

- Daniel está 3 desviaciones estándar arriba de la media en términos de su coeficiente intelectual (CI). Describe su intelecto general.
- Explica por qué es probable que una distribución no sea normal cuando la desviación estándar es más grande que la media.

EJERCICIOS PARA EL CAPÍTULO 5

Conjunto de ejercicios 5A

- 5A-1. Utiliza la fórmula de la desviación estándar para completar los espacios en blanco de la tabla siguiente. La tabla presenta cálculos en variables de intervalo/razón para diferentes muestras de tamaño n .

Suma de cuadrados	n	Varianza	Desviación estándar
11 828.52	88	135.96	_____
3 120.00	21	_____	_____
893.49	_____	30.81	_____
_____	347	124.65	11.16

- 5A-2. Hughes y Waite (2002) estudiaron las condiciones de vida y salud en los años tardíos de la edad media del ciclo vital. Supongamos que la siguiente es una serie de edades del estudio que realizaron estos investigadores: 74, 81, 83, 77, 76, 79, 79.
- Organiza una hoja de cálculo con casos ordenados con X = edad.
 - Calcula la edad media, mediana y modal.
 - Calcula el rango de las edades.
 - Calcula la desviación estándar de las edades.
- 5A-3. Hoff (2003) examinó las vidas laborales de médicos empleados por organizaciones de conservación de la salud (OCS). Supongamos que los datos siguientes describen la carga diaria de pacientes (es decir, el número de pacientes revisados por día) de siete médicos de las OCS: 8, 7, 11, 4, 5, 13, 7.
- Organiza una hoja de cálculo con casos ordenados con Y = número de pacientes revisados por día.
 - Calcula la media, mediana y modal de las cargas de pacientes para este grupo de siete médicos.
 - Calcula el rango de pacientes revisados por día.
 - Calcula la desviación estándar.
- 5A-4. Takao y otros (2003) examinaron la relación entre grupo ocupacional y actividad física entre empleados japoneses. Supongamos que los datos siguientes representan una muestra de puntuaciones en una escala que mide las posiciones de individuos

dentro de la jerarquía ocupacional japonesa: 27, 26, 28, 30, 31, 29, 27, 31, 29. Los datos X = puntuación de escala de grupo ocupacional.

- Organiza los datos usando un formato de hoja de cálculo con las puntuaciones de X ordenadas.
- Calcula la media y desviación estándar.

5A-5. Es semana de regreso a casa y están ocurriendo locuras en el *campus*. Una de estas locuras es una carrera a pie entre el club femenino de estudiantes. Una muestra aleatoria de hermanas que corren, prendas de garantía y alumnas produce las edades siguientes: 19, 18, 20, 19, 29, 18, 20, 18, 22, 21. Los datos X = edad.

- Organiza los datos usando un formato de hoja de cálculo con las puntuaciones de X ordenadas.
- Calcula la media y desviación estándar.
- ¿Hay algo peculiar en esta distribución? Ajústalo al recaular los estadísticos.
- Comenta sobre las diferencias entre los estadísticos original y ajustado.

5A-6. Ellickson y cols. (2003) examinaron la conducta de fumar en adolescentes y la subsecuente conducta después de fumar. Supongamos que los siguientes datos son de una muestra de fumadores de 16 a 20 años de edad.

Y = número de cigarrillos fumados por día.

$\bar{Y}$ = 15 cigarrillos s_y = 5 cigarrillos

- Completa las columnas de la tabla siguiente. Asegúrate de especificar las unidades de medida.
- ¿Quién destaca como fumador?

Caso	Y (cigarrillos por día)	$Y - \bar{Y}$ (desviación estándar)	Z_y (puntuación estandarizada)
Bob Smith	17		
Spencer Byrd	30		
Sonya Turnham	4		
Chuck Martin	20		

5A-7. Ferraro y Yu (1995) estudiaron la relación entre el peso corporal y la autoestima en cuanto a salud. Supongamos que te dan los siguientes breves estadísticos acerca del peso obtenido en este estudio.

X = peso $\bar{X}$ = 169 libras s_x = 18 libras

- Traza la curva normal y ubica estos pesos en ésta.
- La tabla siguiente incluye datos para unas pocas de las observaciones. Completa la columna central al estimar cada una de las Z_x visuales (es decir, con sólo observar X en la curva).
- Para cada puntuación X , calcula la puntuación Z exacta e insértala en la columna de la derecha. (Demuestra la fórmula y cálculo para X = 128 libras.)

X (libras)	Estimación visual de puntuación Z (SD)	Puntuación Z calculada (SD)
169		
128		
192		
177		
151		
109		

Conjunto de problemas 5B

5B-1. Usa la fórmula para la desviación estándar para completar los espacios en blanco de la tabla siguiente. La tabla presenta cálculos sobre variables de intervalo/razón de diferentes muestras de tamaño n .

Suma de cuadrados	n	Varianza	Desviación estándar
38.76	7		
347 295.92	1041		18.27
	91	40.89	
5865.04		17.56	

5B-2. Goessling (2001) examinó el fenómeno de la desigualdad de ingresos en todo el mundo, tanto dentro como entre las naciones. Supongamos que lo siguiente es una muestra de ingresos mensuales de residentes de Estados Unidos: \$2 347; \$2 434; \$1 636; \$1 963; \$2 358; \$1 968; \$2 683.

- Organiza una hoja de cálculo con los casos en orden con X = ingreso mensual.
- Calcula el ingreso mensual medio, mediano y modal.
- Calcula el rango.
- Calcula la desviación estándar.

5B-3. Wiesner (2003) examinó relaciones recíprocas entre síntomas depresivos y conducta delictiva entre adolescentes hombres y mujeres. Supongamos que las siguientes son las edades de adolescentes comprendidos en este estudio: 10, 8, 9, 11, 12, 9, 13.

- Organiza una hoja de cálculo con los casos en orden con Y = edad.
- Calcula las edades media, mediana y modal para los adolescentes seleccionados.
- Calcula el rango de edades.
- Calcula la desviación estándar.

5B-4. Groome y Soureti (2004) estudiaron la relación entre el desorden de estrés posttraumático y síntomas de ansiedad en niños, después del terremoto de 1999 cerca de Atenas, Grecia. Entre los datos que presentan en su estudio analizan la magnitud de diversos terremotos en la escala Richter ocurridos en el mar Mediterráneo a fines

del siglo veinte: 5.8, 2.4, 2.2, 6.0, 3.1, 2.4, 2.2, 5.8, 2.4. Los datos X = puntos de magnitud en la escala de Richter.

- Organiza los datos usando un formato de hoja de cálculo con puntuaciones X ordenadas.
- Calcula la media y desviación estándar.

5B-5. Betts y Morell (1999) analizaron los efectos de los antecedentes personales (es decir, de familia, de secundaria, recursos, grupo paritario, etc.) sobre el promedio de puntos de calificación de pasantes (GPA). Supongamos que las siguientes fueron puntuaciones GPA para una muestra de estudiantes universitarios pasantes: 3.6, 3.8, 3.6, 3.9, 2.6, 3.8, 3.8, 3.9. Los datos X = GPA.

- Organiza los datos usando un formato de hoja de cálculo con puntuaciones ordenadas.
- Calcula la media y desviación estándar.
- ¿Hay algo peculiar en esta distribución? Ajústalo al recalcular los estadísticos.
- Comenta sobre las diferencias entre los estadísticos original y ajustado.

5B-6. Green y cols. (2001) estudiaron el fenómeno de delitos con violencia y discutieron las dificultades prácticas asociadas con la recolección de datos sobre este fenómeno. No obstante, supongamos que tú has podido asegurar datos confiables acerca de delitos con violencia en Estados Unidos. Los datos Y = porcentaje de delitos con violencia = número de delitos con violencia reportados por 100 000 habitantes cubiertos por agencias periodísticas. Los siguientes son porcentajes para una muestra seleccionada de estados:

$$\bar{Y} = \text{delitos con violencia por 100 000 habitantes}$$

$$s_y = .32 \text{ delitos con violencia por 100 000 habitantes}$$

- Completa las columnas de la tabla siguiente. Asegúrate de especificar las unidades de medida.
- ¿Qué resalta por tener un porcentaje relativamente alto de delitos con violencia?

Estado	Y (porcentaje de delitos con violencia)	$Y - \bar{Y}$ (puntuación de desviación)	Z , (puntuación estandarizada)
Florida	1.15		
Indiana	1.08		
Iowa	1.02		
Mississippi	.97		
Texas	1.75		

5B-7. Slater y cols. (2003) examinaron la relación entre el contenido violento de medios masivos de comunicación y la conducta agresiva entre adolescentes. Para replicar los resultados obtenidos por estos investigadores, supongamos que llevamos a cabo un análisis de conducta agresiva similar entre muchachos de 13 a 16 años de un centro de detención juvenil. La variable se operacionaliza como el número de actos agresivos, es decir, insultos verbales y amenazas, actos de violencia física y destruc-

ción de propiedades cometidos en la semana previa. Los actos se tabulan al observar videocintas de salas y secciones de recreación, biblioteca, sanitarios y cafetería del centro de detención. Calculamos los estadísticos descriptivos sobre esta variable y obtenemos los siguientes resultados con X = número de actos agresivos:

$$\bar{X} = 16.8 \text{ actos} \quad s_x = 4.4 \text{ actos}$$

- Traza la curva normal y ubica estos actos agresivos en la curva.
- La tabla siguiente incluye información para unas pocas observaciones. Completa la columna central al estimar cada Z_x visual (es decir, con sólo observar X en la curva).
- Para cada puntuación X , calcula la puntuación Z exacta e insértala en la columna derecha. (Muestra la fórmula y cálculo para $X = 9$ actos.)

X (actos agresivos)	Estimación visual de puntuación Z (SD)	Puntuación Z calculada (SD)
9		
12		
19		
26		
3		
14		

Conjunto de problemas 5C

5C-1. Usa la fórmula para la desviación estándar para completar los espacios en blanco de la tabla siguiente. La tabla presenta los cálculos para variables de intervalo/razón de diferentes muestras de tamaño n .

Suma de cuadrados	n	Varianza	Desviación estándar
12654.27	97	131.82	
2876.54	18		
975.46		34.82	
	526	142.53	

5C-2. Para los sujetos de investigación en comunidades de bajos ingresos en Kenya, Molyneux y otros (2004) examinaron la comprensión de documentos de un acuerdo informado. Como parte del componente cuantitativo del estudio, supongamos que los siguientes son ingresos de residentes que viven dentro de la zona de investigación de los autores: \$ 627, \$435, \$569, \$615, \$796, \$715, \$615.

- Organiza los datos usando un formato de hoja de cálculo con casos ordenados con X = ingresos.
- Calcula el ingreso medio, mediano y modal.
- Calcula el rango de ingresos.
- Calcula la desviación estándar de ingresos.

- 5C-3. Siebert (2004) estudió los determinantes de la depresión entre trabajadoras sociales en Carolina del Norte. Supongamos que los datos siguientes representan el número de contactos en la semana pasada para cada trabajadora social de la muestra: 10, 8, 13, 7, 6, 15, 6.
- Organiza una hoja de cálculo con los casos en orden con Y = contactos en la semana pasada.
 - Calcula la media, mediana y moda del número de contactos para la muestra de trabajadoras sociales.
 - Calcula el rango.
 - Calcula la desviación estándar.
- 5C-4. Roose y cols. (2004) trataron de determinar la eficacia de medicamentos para el tratamiento de síntomas depresivos en pacientes de 75 años de edad o más. Supongamos que los datos siguientes son puntuaciones de la escala de depresión del Centro de Estudios Epidemiológicos (CESD), para un pequeño número de participantes de estudio: 38, 31, 42, 27, 19, 49, 31, 19, 38. Los datos X = puntos de la escala CESD.
- Organiza los datos usando un formato de hoja de cálculo con puntuaciones ordenadas.
 - Calcula la media y desviación estándar.
- 5C-5. Ebrahim y cols. (2004) estimaron la asociación entre la posición socioeconómica y la incapacidad autorreportada en hombres de edad avanzada. Supongamos que las siguientes son las edades de varios hombres en este estudio: 74, 69, 76, 72, 72, 78, 87, 74, 69, 74. Los datos X = edad.
- Organiza los datos usando un formato de hoja de cálculo con puntuaciones ordenadas.
 - Calcula la media y desviación estándar.
 - ¿Hay algo peculiar en esta distribución? Ajústalo al recalcular los estadísticos.
 - Comenta sobre las diferencias entre los estadísticos original y ajustado.
- 5C-6. Ramstedt (2004) examinó el consumo de alcohol y la mortalidad relacionada con el alcohol en Canadá. Supongamos que los datos siguientes son de una submuestra de sujetos adictos al alcohol, tomados de la investigación del autor. Los datos Y = número de copas de alcohol ingeridas por día.

$$\bar{Y} = 6 \text{ copas}$$

$$s_Y = 2 \text{ copas}$$

- Completa las columnas de la tabla siguiente. Asegúrate de especificar las unidades de medida.
- ¿Quién destaca como el más bebedor?

Sujeto	Y (copas por día)	$Y - \bar{Y}$ (puntuación de desviación)	Z_Y (puntuación estandarizada)
Jill Williams	4		
Thomas Wilke	8		
Jason Schmidt	12		
Jenny Pence	7		

- 5C-7. El índice de masa corporal (IMC) es una medida del nivel saludable de peso que toma en cuenta la estatura de personas. Se calcula en kilogramos de peso por el cuadrado de la estatura. Xiaoxing y Baker (2004) investigaron la relación entre el IMC, la actividad física y el riesgo de ambos en la disminución de la salud general y el funcionamiento físico. Supongamos que te dan los siguientes y breves estadísticos sobre el IMC de este estudio. Los datos X = puntuación IMC.

$$\bar{X} = 31 \text{ kg/m}^2$$

$$s_X = 9 \text{ kg/m}^2$$

- Traza la curva normal y ubica estos índices de masa corporal en la curva.
- La tabla siguiente incluye datos para algunas de las observaciones. Completa la columna central al estimar cada una de las Z_X visuales (es decir, con sólo observar X en la curva).
- Para cada puntuación X , calcula la puntuación Z exacta e insértala en la columna de la derecha. (Muestra la fórmula y cálculo para $X = 11$.)

X (libras)	Estimación visual de puntuación Z (SD)	Puntuación Z calculada (SD)
31		
11		
42		
38		
22		
2		

Conjunto de problemas 5D

- 5D-1. Usa la fórmula de la desviación estándar para completar los espacios en blanco de la tabla siguiente. La tabla presenta los cálculos sobre variables de intervalo/razón de diferentes muestras de tamaño n .

Suma de cuadrados	n	Varianza	Desviación estándar
29.57	5		
426 113.21	1986		14.65
	82	35.43	
8450.35		22.12	

- 5D-2. Garroute y cols. (2004) examinaron la relación entre la identidad étnica y la satisfacción entre pacientes indios norteamericanos crónicamente enfermos, en una clí-

nica de la nación Cherokee. Supongamos que la siguiente es una serie de edades del estudio de estos investigadores: 56, 64, 62, 57, 64, 59, 58.

- Organiza una hoja de cálculo con los casos en orden con X = edad.
- Calcula la media, mediana y moda para la edad.
- Calcula el rango de edades.
- Calcula la desviación estándar de las edades.

5D-3. Henning y Feder (2004) compararon las diversas características de hombres y mujeres encarcelados por violencia doméstica. Supongamos que la muestra siguiente representa el número de incidentes informados de violencia doméstica entre los hombres y mujeres incluidos en este estudio: 5, 2, 3, 4, 3, 6, 1.

- Organiza una hoja de cálculo con los casos en orden con Y = incidentes de violencia doméstica.
- Calcula la media, mediana y moda de los incidentes.
- Calcula el rango.
- Calcula la desviación estándar.

5D-4. Tohill y cols. (2004) revisaron la evidencia epidemiológica que evalúa la relación entre consumo de frutas y verduras y el peso corporal. Supongamos que los siguientes son pesos corporales promedio de un pequeño número de participantes: 168, 181, 144, 159, 181, 204, 168, 144, 181. Donde X = peso corporal (libras).

- Organiza los datos usando un formato de hoja de cálculo con las puntuaciones ordenadas.
- Calcula la media y desviación estándar.

5D-5. Grove y Wasserman (2004) examinaron patrones de ciclos de vida del promedio de puntos de calificaciones (GPA) entre cinco grupos colegiados en una importante universidad privada. Supongamos que las siguientes son una muestra de las GPA colegiadas de varios sujetos de estudio en esta investigación: 3.8, 3.4, 2.4, 3.8, 3.7, 3.4, 3.4, 3.8. Donde X = GPA.

- Organiza los datos usando un formato de hoja de cálculo con las puntuaciones ordenadas.
- Calcula la media y desviación estándar.
- ¿Hay algo peculiar en esta distribución? Ajústalo al recalcular los estadísticos.
- Comenta sobre las diferencias entre los estadísticos original y ajustado.

5D-6. Varano y cols. (2004) exploraron la correlación entre el consumo de drogas y homicidios. Supongamos que los estadísticos siguientes representan los porcentajes de homicidios de las zonas de investigación incluidas en este análisis. Donde Y = porcentaje de homicidios = número de homicidios informados por 100 000 habitantes.

$$\bar{Y} = 6.59 \text{ homicidios por } 100\,000 \text{ habitantes}$$

$$s_Y = 1.74 \text{ homicidios por } 100\,000 \text{ habitantes}$$

- Completa las columnas de la tabla siguiente. Asegúrate de especificar las unidades de medida.
- ¿Qué zona destaca por tener un porcentaje relativamente alto de homicidios?

Estado	Y (porcentaje de homicidios)	$Y - \bar{Y}$ (puntuación de desviación)	Z_Y (puntuación estandarizada)
Área 1	4.97		
Área 2	8.99		
Área 3	5.99		
Área 4	6.95		
Área 5	6.29		

5D-7. Boardman (2004) evaluó la relación entre la estabilidad residencial y la salud física entre adultos de raza negra y blanca. Parte de la variación en los niveles de salud se debía a diferencias en los niveles de estrés entre vecindarios. Supongamos que te dan los siguientes estadísticos de resumen de una escala empleada para analizar niveles de estrés en vecindarios. Donde X = puntuación en la escala de estrés.

$$\bar{X} = 11.3 \text{ puntos de la escala de estrés}$$

$$s_X = 3.2 \text{ puntos de la escala de estrés}$$

- Traza la curva normal para estas puntuaciones de escala de estrés y ubícalas en la curva.
- La tabla siguiente incluye datos para algunas de las observaciones. Completa la columna central al estimar cada una de las Z_X visuales (es decir, con sólo observar X en la curva).
- Para cada puntuación X , calcula la puntuación Z exacta e insértala en la columna de la derecha. (Demuestra la fórmula y cálculo para $X = 7$ puntos de escala de estrés.)

X (puntos de la escala de estrés)	Estimación visual de puntuación Z (SD)	Puntuación Z calculada (SD)
7		
10		
14		
19		
1		
16		

APLICACIONES OPCIONALES DE COMPUTADORA PARA EL CAPÍTULO 5

En el sitio web www.mhhe.com/richey2, en *The Statistical Imagination*, están disponibles ejercicios computarizados opcionales del capítulo. Además, el apéndice D de este texto, *Guide to SPSS for Windows*, contiene instrucciones básicas para calcular estadísticos de dispersión y puntuaciones estandarizadas. Tal como se destaca en este capítulo, la desviación estándar por lo general se reporta con la media. De este modo, estos "estadísticos descriptivos" se encuentran juntos en el software. La media, rango y desviación estadística se pueden calcular desde varios lugares del SPSS y son parte de estadísticos opcionales para numerosos procedimientos estadísticos de prueba, como los de los capítulos 9-12.