CAPÍTULO 4

Estimación de promedios

RESUMEN DEL CAPÍTULO

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 $\overline{X}$, 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

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

donde

$\overline{X}$ = 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

$$\overline{X} = \frac{\Sigma X}{n} = \frac{6 + 12 + 5 + 10 + 9 + 10 + 8 + 7 + 9 + 11 + 8 + 10}{12}$$

$$= \frac{105 \text{ años}}{12} = 8.75 \text{ años}$$

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 secretarias de un banco local, y, para el grupo 2, los tres vicepresidentes. Para las ocho secretarias:

$$\overline{X}_{(grupo1)} = \frac{\Sigma X_{(grupo1)}}{n_{(grupo1)}} = \frac{7 + 10 + 7 + 12 + 16 + 7 + 14 + 10}{8}$$

$$= \frac{83 \text{ días}}{8} = 10.38 \text{ días de vacaciones}$$

Para los tres vicepresidentes:

$$\overline{X}_{(grupo2)} = \frac{\Sigma X_{(grupo2)}}{n_{(grupo2)}} = \frac{60 + 30 + 30}{3}$$

$$= \frac{120 \text{ días}}{3} = 40.00 \text{ días de vacaciones}$$

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

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

$$= \frac{83 + 120}{8 + 3} = \frac{203}{11} = 18.45 \text{ días de vacaciones}$$

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:

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

donde

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

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

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:

$$\overline{X}_{(de\ último\ y\ penúltimo)} = \frac{\Sigma X_{(último)} + \Sigma X_{(penúltimo)}}{n_{(último)} + n_{(penúltimo)}}$$

$$= \frac{1\,131 + 1\,264}{13 + 16} = \frac{2\,395}{29} = 82.59 \text{ puntos de examen}$$

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 "encogiendo" 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):

$$\overline{X} = \frac{\Sigma X}{n} = \frac{5 + 2 + 6 + 10 + 8 + 3 + 9 + 11 + 5 + 400}{10}$$

$$= \frac{\$459}{10} = \$45.90 \approx \$46$$

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

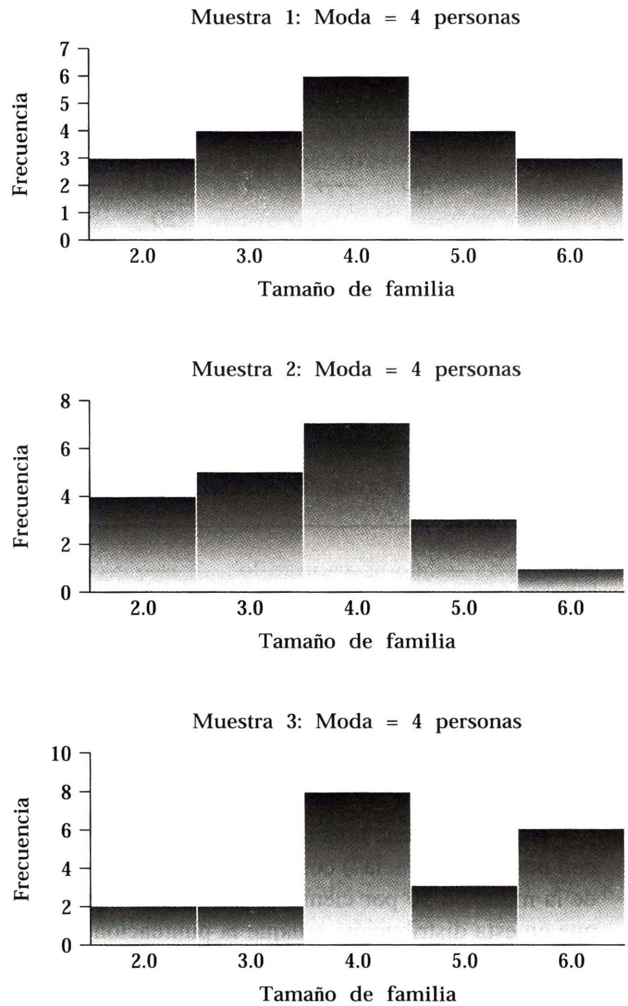


TABLA 4-2 I Distribución de sueldos en un restaurante de comida rápida

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 podríamos 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 la 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.

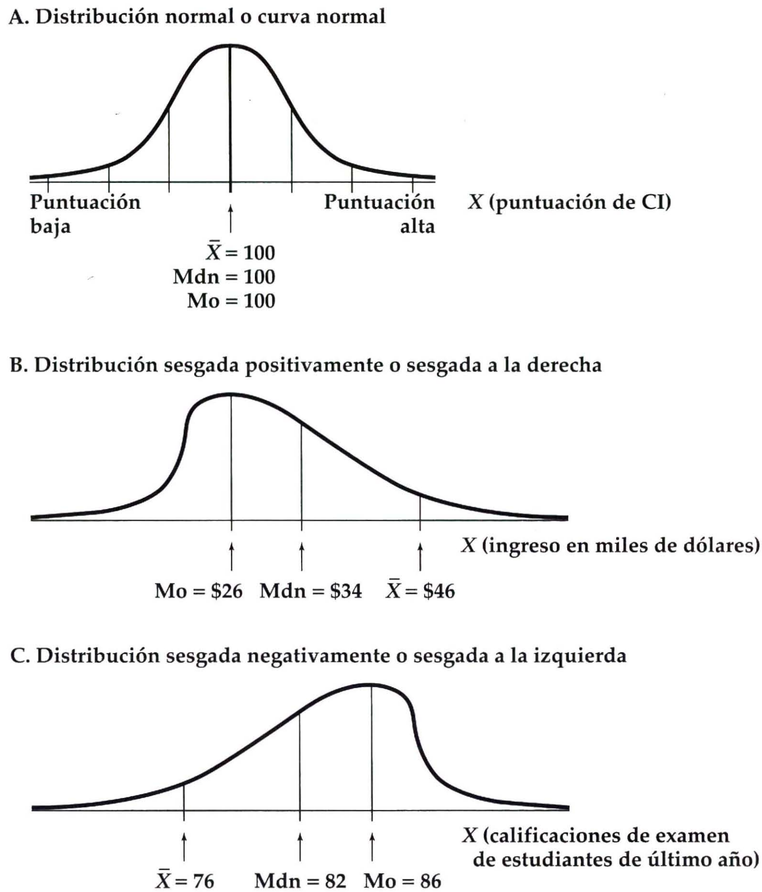


FIGURA 4-2 Curvas de distribución de frecuencias comunes y posiciones relativas de la media, la mediana y la moda, donde X es una variable de intervalo/razón (datos ficticios)

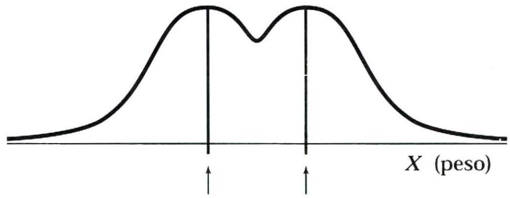
> **Distribución normal** Curva de distribución de frecuencias donde la media, la mediana y la moda de una variable son iguales entre sí y la distribución de las puntuaciones tiene forma de campana.

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 I Características, aspectos fuertes, y debilidades potenciales de la media, mediana y moda

		Aspectos fuertes y aplicaciones		
Estadístico de tendencia central	**Definición**	**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:

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

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
$n = 19$		$\Sigma X = 72$ películas

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

$$\text{Mdn} = 4 \text{ películas}$$

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:

$$\text{Mo} = 4 \text{ películas}$$

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 | 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:

$$\text{Mo} = 4 \text{ películas}$$

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:

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

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

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

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

$$\text{Como } \overline{X} = \frac{\Sigma X}{n},\ \Sigma X = (n)\,(\overline{X})$$

Sustituye para obtener:

$$\overline{X}_{(grupos\ 1\ y\ 2\ combinados)} = \frac{n_{(grupo1)}\overline{X}_{(grupo1)} + n_{(grupo2)}\overline{X}_{(grupo2)}}{n_{(grupo1)} + n_{(grupo2)}}$$

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 Mo = 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.

a) Con esta información, construye las curvas de frecuencias de edad de mortalidad para cada nación.

b) ¿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.

a) Calcula la media y la mediana.

b) En comparación con las otras puntuaciones, ¿cómo llamaríamos al salario de \$30 por hora?

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

d) 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.

a) Calcula la media y la mediana.

b) En comparación con las otras puntuaciones, ¿cómo llamaríamos al precio de colegiatura de \$29 000?

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

d) 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.

a) Organiza los datos en una tabla de hoja de cálculo con las puntuaciones ordenadas bajo una columna marcada "X (ordenada)".

b) Calcula los tiempos de reacción de media, mediana y modal de Flossy.

c) 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	$\overline{X}$	Mdn	Mo	Forma de curva	Trazo 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) A partir de esta información, construye las curvas de frecuencias para cada una de las ciudades.

b) ¿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.

a) Organiza los datos en una tabla de hoja de cálculo con las puntuaciones ordenadas bajo una columna marcada "X (ordenada)".

b) Calcula la media y la mediana de Y.

c) En comparación con las otras puntuaciones, ¿cómo llamaríamos al GPA de 3.9?

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

e) 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.

a) Calcula la media y mediana de calificación del examen.

b) En comparación con las otras puntuaciones, ¿cómo llamaríamos a la puntuación de 64?

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

d) 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.

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

b) Calcula los porcentajes de bateo de media, mediana y moda del equipo.

c) 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	Trazo de la curva
Estatura (pulgadas)	70	68	66		
Exámenes este semestre	10	13	15		
Puntuación de espiritualidad	30	30	30		
Presupuesto de abarrotes	$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 "ganes".

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) A partir de esta información, construye las curvas de frecuencias de las calificaciones del examen para cada grupo.

b) ¿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.

a) Organiza los datos en una tabla de hoja de cálculo con las puntuaciones ordenadas bajo una columna marcada "*Y* (ordenada)".

b) Calcula el salario medio y mediano.

c) En comparación con otras puntuaciones, ¿cómo llamaríamos al salario de $212 000?

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

e) 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.

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 evaluación de 3?

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

e) 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).

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

b) Calcula los tiempos de media, mediana y moda de la carrera de 40 yardas del grupo.

c) 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	$\overline{X}$	Mdn	Mo	Forma de la curva	Trazo 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.

a) De esta información, construye las curvas de frecuencias de ingreso familiar para cada comunidad.

b) ¿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 1 140 puntos GRE. La puntuación media para las 54 solicitantes mujeres es 1 210. ¿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	Trazo 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.