- **Ritchey**, Ferris (2001) *Estadística para las ciencias sociales. El potencial de la imaginación estadística* (México D.F.: McGraw-Hill/Interamericana Editores) Capítulo 3: "Gráficos: una imagen dice más que mil palabras" y Capítulo 4: "Estimación de promedios".

CAPÍTULO 4
ESTIMACIÓN DE PROMEDIOS

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 -altura, peso, inteligencia, etcétera- 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". Llamamos a esta puntuación típica, promedio 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 alturas 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 esa 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 fortalezas, 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 solo 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 las calificaciones en un examen 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. También puede calcularse para variables ordinales de tipo intervalo; pero se debe tener cuidado en la interpretación de los resultados.

En 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 lectora, 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 de la clase de estadística es 20.5 años, decimos, "X barra es igual a 20.5 años". Recuerde especificar las unidades de medición de la variable, en este caso, años. La media se calcula como sigue (Σ se lee como "la suma de").

Cálculo de la media

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

donde

$\bar{X}$ = la media de la variable X de intervalo/razón calculada 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 de la muestra)

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

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

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

años

12

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 en una muestra si cada sujeto de la muestra tuviera la misma puntuación. En el ejemplo anterior, 8.75 años (es decir, ocho 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 cada persona recibiría 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 cual se equilibran las *diferencias entre* la media de X y las puntuaciones individuales X en 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 no informaran sus edades, la suma de edades se dividiría entre 47 -el número de puntuaciones válidas- en lugar de dividirla entre el tamaño de la muestra (49). Es más, con archivos de computadora, debe tenerse cuidado de no sumar los códigos de "valor perdido" (como 99) 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. 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 no es un error es cuando los dos grupos tienen los mismos tamaños de muestra (es decir, cuando las n son iguales).] Por ejemplo, observe 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:

$$\bar{X}_{(\text{grupo } 1)} = \frac{\Sigma X_{(\text{grupo } 1)}}{n_{(\text{grupo } 1)}} = \frac{7 + 10 + 7 + 12 + 16 + 7 + 14 + 10}{8}$$

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

Para los tres vicepresidentes:

$$\bar{X}_{(\text{grupo } 2)} = \frac{\Sigma X_{(\text{grupo } 2)}}{n_{(\text{grupo } 2)}} = \frac{60 + 30 + 30}{3}$$

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

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:

$$\bar{X}_{(\text{grupo 1 más 2 combinados})} = \frac{\Sigma X_{(\text{grupo } 1)} + \Sigma X_{(\text{grupo } 2)}}{n_{(\text{grupo } 1)} + n_{(\text{grupo} 2)}}$$

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

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.

Encerrado en un promedio. Todos encontramos situaciones en las cuales nuestro promedio para un desempeño parece dirigirse a cierto nivel. No importa cuánto mejoremos, el promedio parece estar encerrado. Por ejemplo, la media sirve para calcular promedios en el boliche, las cuales van desde cero hasta 300. Brian practica boliche semanalmente en una liga. Después de 100 juegos su promedio es de 150 pinos por juego. ¿Qué puntuación debe obtener en el próximo juego para subir su promedio 1 punto?

El pensamiento proporcional sobre un "promedio móvil" guía nuestros cálculos de la media. Primero enfoquémonos en el denominador, n. Después del juego 101, habrá aumentado de 100 a 101. El numerador, ΣX, aumentará por la cantidad de la puntuación de Brian en el juego 101. La pregunta es: ¿cuánto debe aumentar ΣX para elevar su media a 151 pinos por juego?

¡Quizás Brian se sienta defraudado al saber que debe alcanzar una puntuación de 251 en el juego 101, para levantar su promedio sólo un punto! ¿Por qué tanto? Determinemos la puntuación total que Brian ha acumulado hasta el juego 100 con su promedio de 150. Esto se calcula resolviendo ΣX en la ecuación para la media:
Puesto que

$$\bar{X} = \frac{\Sigma X}{n} \quad \text{Entonces} \quad \Sigma X = (n)(\bar{X})$$

Puntuación total de Brian después de 100 juegos es:

$\Sigma X_{\text{(hasta el juego 100)}} = (n)\ (\bar{X}) = (100)\ (150) = 15.000$ pinos

Si su promedio debe incrementarse a 151, su ΣX después del juego 101 debe ser

$\Sigma X_{\text{(necesaria después del juego 101)}} = (n)\ (\bar{X}) = (1001)\ (151) = 15.251$ pinos

La puntuación que requiere en el juego 101 es la diferencia entre estas dos sumas de X:
La puntuación que Brian necesita para subir su promedio a 151 después de 101 juegos

$= \Sigma X_{\text{(necesaria después del juego 101)}} - \Sigma X_{\text{(hasta el juego 100)}}$
$= 15.251 - 15.000 = 251$ pinos

Esto significa que para que Brian eleve 1 punto su promedio, debe aumentar 1 punto en el juego 101 *más un punto extra para todos los 100 juegos previos*. De hecho, si sólo restan unas cuantas semanas de juego en la liga, el promedio de Brian está bastante encerrado en este momento.

Con un promedio móvil, el tamaño de la muestra (n) aumenta regularmente con el tiempo en pequeños incrementos, tanto como tres juegos de boliche por semana, o tres turnos al bat por juego para los jugadores de la liga de béisbol profesional. Para elevar un promedio, un jugador necesita igualar su promedio anterior más algún valor. Conforme la temporada continúa, la cantidad de este "más algún" sigue creciendo. De la misma forma, sin embargo, es fácil mejorar un primer promedio en la temporada. Por ejemplo, si Brian obtiene 150 en el primer juego de la temporada, ¿cuánto tendrá que conseguir en el segundo juego para elevar su promedio a 151? Por último, esto resalta un punto importante respecto de la evaluación exacta del desempeño de alguien en un punto determinado: es más apropiado usar la media de juegos recientes en lugar de la de la temporada entera.

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 (aumentar) 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 suponga 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):

$$\bar{X} = \frac{\Sigma X}{n} = \frac{5+26+10+8+2+9+11+5+400}{10}$$

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

Por obvias razones, esta media de $46 no representa la cantidad de dinero promedio típico, o la tendencia central que los alumnos suelen portar 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 extremo. Para obtener un sentido de proporción sobre cómo se calcula la fórmula de la media, examine la relación entre el numerador (ΣX) y el denominador (n). Cuando ΣX es grande y n es pequeña, la media será grande. Si ΣX es grande debido a la presencia de uno o dos valores extremos de alto valor, la media se "inflará" hasta un valor grande.

Tenga presente 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 *en 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, calcule la edad media de la siguiente muestra de cinco estudiantes de la universidad local, donde un estudiante en la muestra tiene una edad extremadamente alta: 19, 19, 20, 21 y 54 años. La respuesta dejará la impresión de que esta muestra está bastante arriba de la edad típica en la universidad, cuando, de hecho, cuatro de los cinco estudiantes *tiene* la edad típica. También observe lo que sucede cuando existe una puntuación sumamente baja, como con esta muestra de edades: 8, 19, 19, 20 y 21 años. En tales casos, los valores extremos deben eliminarse, y la media debe calcularse de nuevo sin ellos. Al informar esta "media ajustada", notamos por qué se realizó el ajuste.

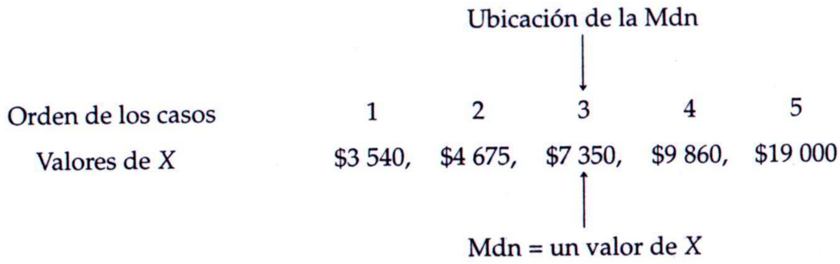
En cualquier momento que calculamos una media, en especial con una muestra pequeña, primero examinamos la distribución de frecuencias de la puntuación en bruto para los valores extremos. Un recurso práctico para esto es un gráfico de caja (capítulo 3). Ya que la media es más útil que la mediana y la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

La mediana

La **mediana** (Mdn) es *la puntuación de la mitad en una distribución ordenada* -aquel valor de una variable que divide la distribución de las puntuaciones por la mitad, *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 de los hogares en la ciudad Combelt es $26.000, la mitad de los hogares en 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 en el que 50 por ciento de las observaciones caen debajo. 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 incrementaría el valor de la media.

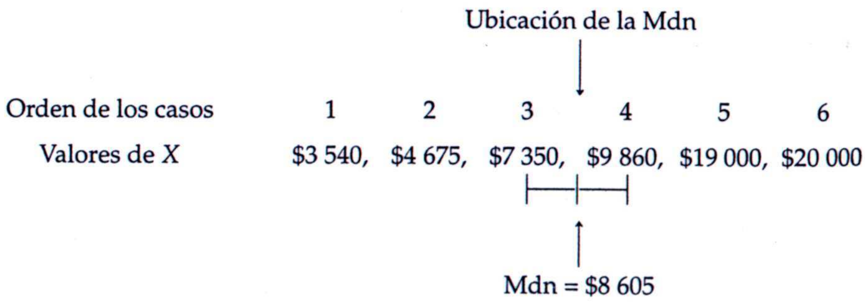
La mediana
Para una variable ordinal o de intervalo/ razón, la puntuación de la mitad en una distribución ordenada, la puntuación por arriba de la cual queda la mitad de los casos y por debajo queda la otra mitad.

Para calcular la mediana de una distribución, primero deben ordenarse 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. Divida el tamaño de la muestra n entre 2 para acercarse a la puntuación de la mitad en la distribución. Si n es un número impar, la mediana será un caso real en la muestra. Suponga, por ejemplo, que tenemos una muestra de cinco familias con los siguientes ingresos mensuales (X):



El ingreso mediano 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 un ingreso de $20.000, se inserta en la muestra anterior,



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 se vuelve una tarea directa. Con una muestra grande (y con la ayuda de un programa de cómputo), la mediana se localiza matemáticamente dividiendo el tamaño de la muestra entre 2 y sumando .5. Observe que este resultado da la *ubicación ordenada* de la mediana, no la mediana en sí. Ordene las puntuaciones, y luego cuente hasta esta posición. La puntuación X en esta posición es la mediana. Después de encontrar la mediana, vuelva a revisar verificando si su respuesta, de hecho, divide los casos por la mitad. La mediana puede usarse con variables de intervalo/ razón, así como con variables ordinales. Finalmente, no confunda la mediana con otro estadístico llamado *rango medio*, que es el punto a la mitad entre los valores mínimo y máximo de X.

Cálculo de la mediana (Mdn)
1. Ordene la distribución de puntuaciones de menor a mayor. **2**. Ubique la posición de la mediana. Divida el tamaño de la muestra, n, entre 2 para ubicarse cerca de la puntuación que está a la mitad en 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 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 el tamaño de la muestra entre 2 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 en una distribución. Es *insensible a los valores de las puntuaciones* en una distribución; es decir, sin tener en cuenta los valores de X que la rodean, la mediana es la puntuación de la mitad determinada por el número de puntuaciones (n) en la muestra. Por ejemplo, las siguientes dos distribuciones de puntuaciones en un examen tienen la misma mediana; aunque 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 en ambas clases es 77 sería impreciso porque sugiere que las dos tuvieron igual desempeño. (De hecho, el aula 2 lo hizo mucho mejor, con una medio de 83.6, comparado con una media de 65.2 para el aula 1.) La mediana no se afecta por los valores de X.

Mientras es insensible para valores de las puntuaciones, la mediana es sensible a (o afectada por) cualquier cambio en el tamaño de la muestra. Por ejemplo, suponga que en el aula 1 dos estudiantes hacen el examen tarde; lo realizan mal, que es típico de estudiantes que llegan tarde a una evaluación. Cuando sus puntuaciones se incluyen en la distribución, la mediana cambia drásticamente de 77 a 51:

Aula 1(incluye las puntuaciones tardías):

34, 36, 39, 51, 77,
78, 81
↑
Mdn

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

La moda

La **moda** (Mo) es *la puntuación que ocurre con mayor frecuencia en una distribución.* Conceptualmente, la moda es la puntuación "más popular". La tabla 4-1 presenta 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. Note que la moda es una puntuación X (19 años), *no* una frecuencia, *f* (49 casos).

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

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

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

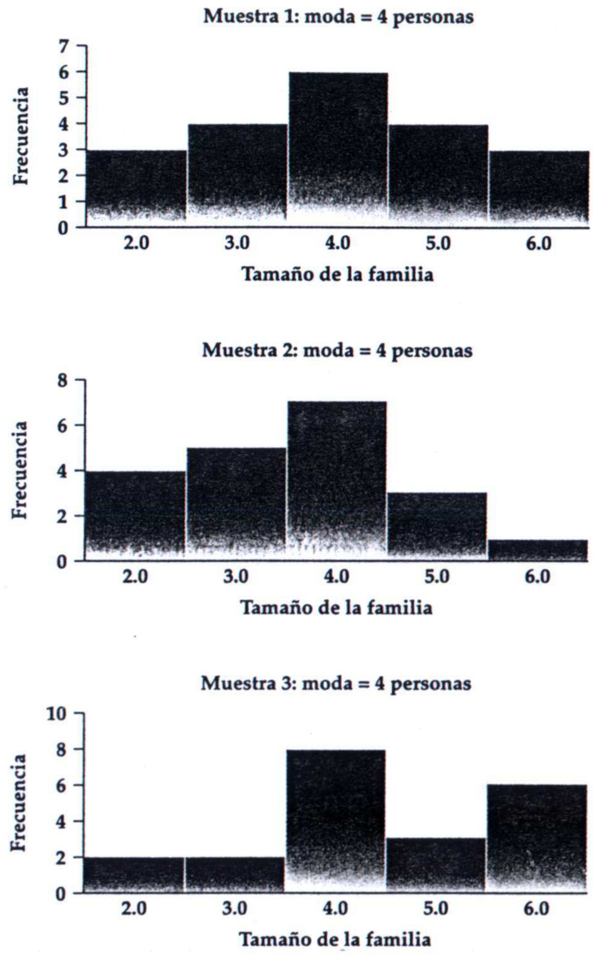
	Especificaciones		**Cálculos**
	Edad	***f***	*Porcentaje*
	18	31	24.8%
Mo→	19	49	39.2
	20	20	16.0
	21	18	14.4
	22	7	5.6
	Total	125	100.0%

Una nota precautoria. No confunda la moda (la "puntuación que ocurre con mayor frecuencia") con la "mayoría de las puntuaciones". Una mayoría simple sería "más de la mitad" 0 50 por ciento de los casos en una muestra más por lo menos, uno. Observe 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 en esta distribución tiene una 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, la barra más alta; en un histograma, la columna más alta; y en un polígono, la puntuación del punto más alto, o el pico.

FIGURA 4-1

Distribuciones de puntuaciones de varias formas con la misma moda



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

En general, *por sí misma* la moda es el estadístico de tendencia central menos útil porque tiene un alcance informativo limitado. Mientras identifica la puntuación que ocurre más frecuentemente, 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 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 en una distribución como al tamaño de la muestra. Esto significa que usted puede tener cualquier número de distribuciones con formas totalmente diferentes, y aun todas podrían tener la misma moda, como se ilustra en la figura 4-1.

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*, 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 informar 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

Recuerde, como vimos en el capítulo 2, que el nivel de medición de una variable nos dice qué fórmulas matemáticas y estadísticos son apropiados para dicha variable. La media y la mediana son claramente apropiadas con variables de intervalo/ razón. Tiene sentido hablar sobre el peso, la altura o el ingreso medios. Tales estadísticos

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

Sueldo ($)	***f***	*Clasificación de empleados*
5.75	12	Empleados regulares
10.50	2	Gerentes nocturnos
18.90	1	Gerente en jefe
Total	16	

también pueden utilizarse con variables ordinales de tipo intervalo, si las frecuencias no se agrupan en un extremo de la distribución, o se dividen entre los dos extremos. Los estadísticos principiantes, sin embargo, deben evitar usar 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 al respecto. Una persona no puede tener un promedio de tanto hombre y tanto mujer; se es uno o lo otro. Recuerde 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 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.* Como 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. Nuestra preocupación está en evaluar la forma de una distribución y examinar las ubicaciones relativas de la media, la mediana y la moda, para estimar la forma de una distribución de frecuencias.*

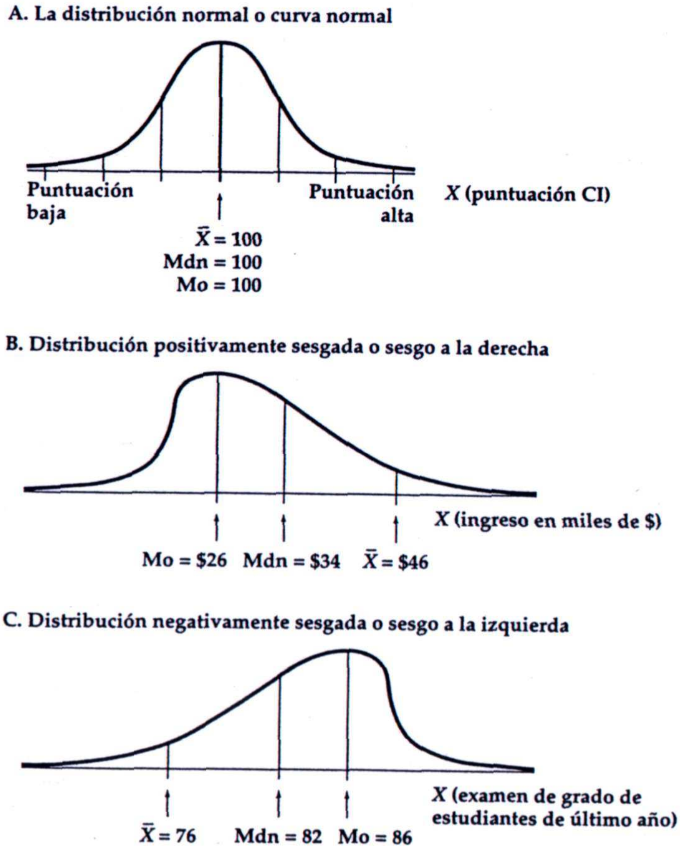
Curva de distribución de frecuencias
Un sustituto de un histograma de frecuencias o polígono donde reemplazamos estos gráficos con una curva suavizada. El área bajo la curva representa el número total de sujetos en la población y es igual a una proporción de 1.00 o a un porcentaje de 100 por ciento.

La figura 4-2 presenta tres formas muy comunes de curvas de distribución de frecuencias de puntuaciones. Como con nuestros histogramas, el eje horizontal de las curvas representa las puntuaciones de una variable X. El eje vertical (el cual a menudo 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 ubicaciones 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 muchos 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 (skewer)* (figura 4-2B y C).

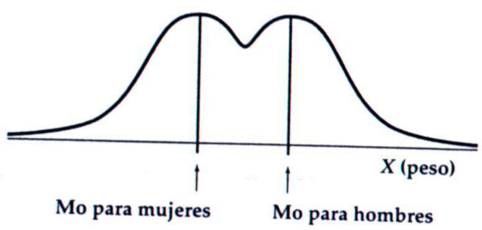
Distribución sesgada Curva de distribución de frecuencias aquella en la cual la media, la mediana y la moda de una variable son desiguales y muchos de los sujetos tienen puntuaciones sumamente altas o bajas.

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 final alto o positivo de la distribución de puntuaciones* (figura 4-28). 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 la moda o, más probablemente, caerá entre éstas.

El **sesgo a la izquierda (o negativo)** tiene *puntuaciones extremas en el final baja 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.

FIGURA 4-3
Distribución bimodal del peso de hombres y mujeres



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

Al calcular 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 de menos) es mayor que 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 en 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 informar 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 una estadística. Sin embargo, si una distribución está sesgada, la mediana es el estadístico que debe 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 una gráfica 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 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. ¿Puede 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, ya que permite mayor flexibilidad en los cálculos matemáticos.

TABLA 4-3 Características, fortalezas y debilidades potenciales de la media, la mediana y la moda

Estadístico de tendencia central	*Definición*	*Fortalezas y aplicaciones*	*Debilidades potenciales*
Media	Valor de X si todas las puntuaciones son las mismas.	Abierta a operaciones matemáticas; preferible cuando la distribución tiene forma normal; útil con variables de intervalo/razón y variables ordinales de tipo intervalo.	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 queda la otra mita.	Preferida cuando la distribución está sesgada; útil con variables de intervalo/razón y variables ordinales de tipo intervalo.	Insensible a los valores de X en la distribución, pero sensible a los cambios en el tamaño de la muestra.
Moda	Puntuación que ocurre con más frecuencia en una distribución.	Preferida cuando virtualmente todas las puntuaciones en la distribución son las mismas, útil con todos los niveles de medición.	Insensible a los valores de X e insensible a cómo se distribuyen a su alrededor las puntuaciones.

En la mayoría de los casos, la mediana y la moda representan callejones sin salida porque no ofrecen operaciones matemáticas adicionales que valgan la pena. Se gana poco con informarlas. Siempre que es posible, la media es la medición sumaria que debe usarse, sobre todo, con estadísticas inferenciales. Debido a esto, a menudo ajustamos distribuciones sesgadas para "hacerlas normales", de manera que podamos usar la media. Se discuten las especificaciones de este tipo de control del error más adelante en este texto. La tabla 4-3 resume las propiedades de los tres estadísticos de tendencia central.

Organización de los datos para calcular estadísticos de tendencia central

Existen dos formatos comunes para organizar datos y calcular estadísticos de tendencia central en tales datos. Un formato es una tabla desglosada de la distribución de puntuaciones en bruto. Como se indicó en el capítulo 2, un formato de tabla desglosada, por lo común se utiliza para la captura de datos en la computadora, pero dichos formatos comúnmente también son usados por las empresas, los gobiernos y grupos comunitarios para mantener los registros de la organización. Los programas de cómputo de hoja de cálculo, como Lotus 1-2-3.

Excel y Corel Quattro Pro, están diseñados especialmente para tal propósito. Estos formatos de hoja de cálculo evolucionaron a partir de la manera lógica como resolvemos problemas a mano -simplemente listando las puntuaciones de una variable en una columna.

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

Formato desglosado para calcular los estadísticos de tendencia central

Suponga que estamos interesados en saber qué tan a menudo los estudiantes de cinematografía en un departamento de comunicaciones de la universidad estudian su disciplina asistiendo a películas de estreno. Recolectamos una muestra aleatoria de 19 estudiantes. Le pedimos a cada uno nombrar las nuevas películas que vieron en cines en el último mes 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 desglosado con los cálculos necesarios para calcular la media. Las puntuaciones se ordenan para facilitar el cálculo de la mediana y la moda.

Primero, calculamos la media:

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

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

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		ΣX = 72 películas

Segundo, calculemos la mediana. Ya ordenamos las puntuaciones, ya que es necesario para calcular la mediana. El tamaño de la muestra (n =19) dividido a la mitad es de aproximadamente nueve casos, y como n es impar, determinamos que el décimo caso es la mediana. En la tabla desglosada contamos hacia abajo al 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 mayor frecuencia es 4:

Mo = 4 películas

Obviamente, emplear una tabla desglosada para hacer cálculos a mano con un gran número de casos sería difícil de manejar. Una manera más fácil de organizar los datos es usar un formato de distribución de frecuencias.

Formato de distribución de frecuencias para calcular estadísticos de tendencia central

La tabla 4-5 presenta los mismos datos de los 19 estudiantes de cinematografía, pero usa un formato de distribución de frecuencias. Trabajando a partir de la tabla desglosada de la tabla 4-4 (como lo haría una computadora), en la tabla 4-5 observamos que hay una frecuencia de tres estudiantes que reportan dos películas, cinco reportan tres películas, y así sucesivamente.

Primero, calculemos la media, para hacerlo, multiplicamos cada puntuación por su frecuencia. Esto es equivalente a sumar las puntuaciones individuales listadas en el formato desglosado de la tabla 4-4. En ambos formatos, la suma de puntuaciones es 72 películas, y así la media resulta 3.79 películas:

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

Segundo, calculemos la mediana. Las puntuaciones en la distribución de frecuencias ya están ordenadas. La posición se determina de la misma manera como en el formato desglosado y sigue siendo el décimo caso. Para localizar el décimo caso, calculamos la frecuencia acumulativa, que es el número de casos en o debajo de una puntuación de la distribución (capítulo 2). Vemos que el décimo caso en la distribución es uno de los seis estudiantes que reportaron cuatro películas. Así:

Mdn = 4 películas

Por último, el cálculo de la moda es bastante fácil con el formato de distribución de frecuencias. En la tabla 4-5 simplemente observamos la columna que lista las frecuencias (es decir, la columna *f*) y notamos qué puntuaciones ocurrieron con la frecuencia más alta. Más estudiantes (seis de ellos) vieron cuatro películas que cualquier otro número de películas durante el último mes:

Mo = 4 películas

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

Especificaciones		Cálculos	
X	*f*	*f*(X)	*f* *acumulativa*
2	3	6	3
3	5	15	8
4	6	24	14
5	3	15	17
6	2	12	19
n = 19		Σ*f*(X) = 72 películas	

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 claramente qué casos o qué sujetos se incluyen en su cálculo. Organizaciones tales como empresas e instituciones escolares, intencionalmente o no, por lo común informan medias que son irreales. Por ejemplo, el vocero de un distrito escolar público puede informar que el sueldo medio de sus maestros es $45.000. Cuando esto ocurra, los maestros probablemente se reunirán en el aula de descanso de la facultad y se preguntarán 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", incluyendo al personal de mayor salario -como consejeros académicos, auxiliares de los directores y directores- todos ellos están certificados como docentes pero rara vez dan clases. Estos administradores quizá hayan sido incluidos porque el "estadístico" 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 sueldos sesgaron la media. Para evitar tal insensatez estadística, deben informarse por separado las medias para subgrupos distintos.

Mezclar rangos de estatus en ocasiones resulta 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 en la compañía gana un sueldo cercano a esa cantidad.

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

Fórmulas en el capítulo 4

Cálculo de la media:
Para trabajar con una tabla desglosada:

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

Para trabajar con una distribución de frecuencias:

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

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

$$\bar{X}_{\text{(grupo 1 más 2 combinados)}} = \frac{\Sigma X_{\text{(grupo 1)}} + \Sigma X_{\text{(grupo 2)}}}{n_{\text{(grupo 1)}} + n_{\text{(grupo 2)}}}$$

Cálculo de la media combinada A-4

Ya que $\bar{X} = \frac{\Sigma X}{n}$ $\quad \Sigma X = (n)\,(\bar{X})$

Sustituya para obtener:

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

Cálculo de la mediana:

1. Ordene la distribución de puntuaciones de menor a mayor.
2. Localice la posición de la mediana. Divida el tamaño de la muestra, n, entre 2 para obtener la puntuación cercana a la puntuación de la mitad en la distribución. (Si trabaja con una distribución de frecuencias, calcule la frecuencia acumulativa para calcular la ubicación de la mediana.) Si n es un número impar, la mediana será un caso real en la muestra. Si n es un número par, la mediana se ubicará entre las dos puntuaciones de la mitad y se calculará tomando la media de estas dos puntuaciones. (Matemáticamente, la posición de la mediana se encuentra dividiendo el tamaño de la muestra entre 2 y sumando .5.)

Cálculo de la moda:

1. Recabe las puntuaciones en una tabla desglosada de puntuaciones brutas ordenadas o en formato de distribución de frecuencias.
2. Identifique la moda (Mo), que es el valor de X con la frecuencia mayor.

Preguntas para el capítulo 4

1. Para cada estadístico de tendencia central, ¿qué nivel de medición de variables es apropiado?
2. Defina la media, la mediana y la moda. Especifique las limitaciones potenciales de cada una.
3. ¿Por qué es mejor calcular las tres mediciones -la media, la mediana y la moda- que confiar sólo en una?
4. Como regla general, es incorrecto calcular la media para dos grupos combinados dividiendo simplemente sus medias separados entre 2. ¿Cuál es la excepción a esta regla?
5. Si una distribución de puntuaciones está sesgada, ¿qué estadístico de tendencia central es el más apropiado para presentarse al público en general? ¿Por qué?
6. En general, la moda de una distribución es el estadístico de tendencia central menos útil. ¿Bajo qué circunstancias, sin embargo, es el estadístico de tendencia central que informa de manera más apropiada?
7. Si la edad modal de una distribución es de 22 años, ¿significa que la mayoría de las personas en esta población tiene 22 años? Explique.

8. ¿Cómo se localiza la moda en un histograma, un polígono y una curva de distribución de frecuencias?

9. ¿Qué representan los ejes horizontales y verticales en una curva de distribución de frecuencias?

10. Describa las características de una curva normal de distribución de frecuencias.

11. Explique en términos generales cómo un sesgo a la izquierda en una distribución de frecuencias afecta los tres promedios comunes: media, mediana y moda.

12. Explique en términos generales cómo un sesgo a la derecha en una distribución de frecuencias afecta los tres promedios comunes: media, mediana y moda.

13. Suponga 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 para esta variable?

14. La puntuación más alta que una persona puede obtener en el boliche es 300. Fred ha llevado su promedio durante 310 juegos y actualmente tiene 179. Sin hacer cálculos, explique por qué es imposible para Fred subir su promedio a 180 en un solo juego.

15. Como se ilustró en "Insensatez y falacias estadísticas" en este capítulo, la media de una variable puede ser una pobre medida de tendencia central cuando existe una mezcla de rangos de estatus dentro de una población. Proporcione un ejemplo de cómo la mezcla de rangos de estatus resulta en una media que no entra en ningún rango.

Ejercicios para el capítulo 4

Recuerde incluir la fórmula, estipular las unidades de medición y contestar la pregunta.

1. Dado lo siguiente, calcule la edad modal, la edad mediana y la edad media. X = edad.

X	Y
14	14
15	17
19	19
19	22
22	28

2. Los siguientes datos son para la variable Y = distancia del centro de trabajo (en millas) para los empleados de un comercio de máquinas copiadoras. Calcule la moda, la mediana y la media de Y.

Y	Y
13	10
9	11
6	14

3	5
12	7

3. Ese individuo fastidioso que anda con patines en línea alrededor del estacionamiento también participa en una competencia de patines en línea. En las últimas 11 carreras ha quedado en los lugares 2, 4,1, 5, 3, 3,1, 3, 2, 3 y 4.

a) Encuentre la moda, la mediana y la media de su posición final típica.
b) Aunque la posición final es una variable ordinal, ¿por qué es razonable calcular estadísticos de tendencia central en ésta?

4. El equipo de boliche de la universidad participó en ocho competencias el año pasado con las siguientes posiciones finales: 4, 3, 2, 2, 3, 1, 2, 1.

a) Encuentre la moda, la mediana y la media de la posición final del equipo.
b) Aunque la posición final es una variable ordinal, ¿por qué es razonable calcular estadísticos de tendencia central en ésta?

5. Siete oficinistas entraron a una competencia para perder peso. Después de unas cuantas semanas de dieta sus pérdidas de peso (en libras) fueron como sigue: 5, 7, 3, 0, 2, 4 y 3. Calcule la pérdida de peso modal, mediana y media.

6. Las puntuaciones en la sección analítica del examen GRE (Graduate Records Examination) de cinco candidatos a un programa para graduados fueron los siguientes: 700, 625, 640, 590 600. Calcule las puntuaciones media y mediana.

7. Paul conduce muy rápido. El es famoso por su velocidad incluso en paseos cortos y gasta cerca de la mitad de su ingreso en neumáticos y frenos. Por lo menos en viajes cortos, su hermano, Terry, deambula lentamente. Ambos salen de casa de su madre conduciendo a un bar local a sólo dos millas a través de las calles de la ciudad. A Terry le toma cuatro minutos llegar allí.

a) En millas por hora (mph), ¿cuál fue el promedio de velocidad de Terry durante el trayecto de dos millas?
b) ¿Qué tan rápido tendría que manejar Paul para llegar al bar en dos minutos y "ahorrar tiempo" en comparación con Terry?
c) ¿Tiene sentido intentar ahorrar tiempo conduciendo deprisa en viajes cortos?

8. Al llegar al juego de liga de esta semana, Lisa ha jugado 56 juegos de boliche y tiene un promedio global de 163 pinos por juego.

a) ¿Cuánto tendría Lisa que promediar esta noche en sus tres juegos para incrementar su promedio global a 165?
b) Dado su desempeño anterior, ¿es probable que Lisa incremente su promedio global en dos puntos esta noche?

9. Los demógrafos estudian las poblaciones de varios estados, comunidades y países. Un asunto de interés es el crecimiento o disminución en el tamaño de una población, la cual es afectada por la rapidez de los nacimientos, cuánto tiempo viven (es decir, longevidad) y a qué edades comúnmente mueren. Una variable es la edad de mortalidad (es decir, edad de muerte). Suponga que en la nación A, la edad modal de mortalidad es 55; la mediana, 60; y la media, 65. En la nación B, la media también es 65; pero la moda es 75, y la mediana, 70.

a) A partir de esta información, construya las curvas de frecuencia para cada nación.
b) ¿Qué nación parece mejor en términos de longevidad?

10. Al evaluar las tasas de delito entre dos ciudades un criminólogo calcula que X = número promedio de vehículos robados por día (durante 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 también es 25, pero la moda es 35, y la mediana, es 30.

a) A partir de esta información, construya curvas de frecuencia para cada ciudad.
b) ¿En qué ciudad se sentiría más seguro al estacionar su automóvil en la calle?

11. Los cinco miembros de una familia trabajan. Sus sueldos por hora son $30, $10.50, $5.15, $12 y $6.

a) Calcule la media y la mediana.
b) Comparado con las otras puntuaciones, ¿cómo llamaríamos a la tasa de $30?
c) ¿Cuál es su efecto en el cálculo de la media?
d) Realice un ajuste para esta peculiaridad volviendo a calcular la media sin ésta.

12. Los siguientes son promedios (PROM) de estudiantes en un programa tutelar: 1.7, 2.6, 2.3, 3.9, 2.2, 1.9, 2.1. Sea Y = PROM.

a) Calcule la media y la mediana.
b) Comparada con las otras calificaciones, ¿cómo llamaríamos a un promedio de 3.9?
c) ¿Cuál es su efecto en el cálculo de la media?
d) Realice un ajuste para esta peculiaridad volviendo a calcular la media sin ésta.

13. La edad media de 47 hombres en el Club de Bridge Sparkesville es 54.8 años. La edad media de las 62 mujeres en el club es 56.4 años. ¿Cuál es la edad media de los 109 miembros?

14. Las siguientes son las edades medias de pacientes con adicción a sustancias en un centro de tratamiento local, clasificados de acuerdo con el tipo de adicción primaria (datos ficticios). Calcule la edad media de todos los pacientes con adicción a sustancias en el centro.

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

15. En un experimento para observar si los pollos pueden distinguir colores, se les premia con granos de maíz cuando uno de ellos pica una almohadilla con el color correcto. Los tiempos de reacción se miden a la centésima de segundo más cercana. 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.

a) Organice los datos en una tabla de distribución de frecuencias.
b) Calcule los tiempos de reacción medio, mediano y modal de Flossy.
c) Describa la forma de la distribución de tiempos de reacción de Flossy.

16. Los promedios de bateo del equipo de béisbol infantil, Fastball Dodgers, son como siguen: .360, .200, .350, .355, .230, .345, .360, .380 y .400.

a) Organice los datos en una tabla de distribución de frecuencias,
b) Calcule los promedios de bateo medio, mediano y modal del equipo.
c) Describa la forma de la distribución.

17. Dados los siguientes estadísticos y lo que sabemos sobre cómo se relacionan dentro de una distribución de puntuaciones, indique la forma de la distribución para cada variable listada.

Variable	$\bar{X}$	*Mdn*	*Mo*	*Forma de curva*
Edad (años)	30	35	39	
Tamaño de la familia	1.1	3.0	2.0	
Años de empleado	11	8	7	
Peso (en libras)	160	132	134	

18. Dados los siguientes estadísticos y lo que sabemos sobre cómo se relacionan dentro de una distribución de puntuaciones, indique la forma de la distribución para cada variable listada.

Variable	$\bar{X}$	*Mdn*	*Mo*	*Forma de curva*
Altura (pulgadas)	70	68	66	
Exámenes este semestre	10	43	15	
Puntuación de espiritualidad	30	30	30	

Presupuesto comestibles	para	\$130	\$109	\$104

Aplicaciones opcionales en computadora para el capítulo 4

Los ejercicios opcionales en computadora para el capítulo 4 se encuentran en el disco compacto *Computer Applications for the Statistical Imagination*. Estos ejercicios incluyen la generación de estadísticos de tendencia central utilizando el *SPSS para Windows* y estadísticos de tendencia central para obtener un sentido de proporción respecto de las formas de distribuciones de puntuaciones.