

Tablas y gráficas: una imagen dice más que mil palabras

RESUMEN DEL CAPÍTULO

Introducción: representación gráfica de datos 78	Graficación de variables de intervalo/razón 86
Lineamientos para graficar 79	Histogramas 86
Graficación de datos nominales/ordinales 80	Polígonos y gráficos de líneas 89
Gráficos de pastel 80	Uso de gráficos en la estadística inferencial y su aplicación en la investigación 93
Gráficos de barras 83	Insensatez y falacias estadísticas: distorsión gráfica 94

Introducción: representación gráfica de datos

En tiempos recientes, el análisis estadístico se refiere a simplificar y resumir grandes cantidades de información. Si el lector es un científico o sólo es un observador informal de lo que pasa a su alrededor, es probable que sus intereses sean entender el comportamiento de una gran población de sujetos u objetos. Por ejemplo, recuerda tu primer día en el campus. Es probable que hayan surgido muchas preguntas en tu mente cuando tratabas de ajustarte a todas las nuevas actividades relacionadas con este interesante entorno. Quizá te preguntaste: ¿qué tan grande es el campus? ¿Cuántos estudiantes están inscritos aquí? ¿Hay muchos de mi pueblo o que estudian una especialización en mi materia? ¿En lo que se refiere a mi capacidad para tener éxito al menos soy un estudiante promedio? ¿Es éste un lugar en donde me sentiré cómodo?

Es natural que los seres humanos deseen simplificar y organizar sus percepciones del mundo que les rodea con sólo resumirlas. Para orientarnos a una nueva situación, rápidamente buscamos nuevas generalizaciones que describen esta gran imagen. Deseamos simplificar las complejidades que nos asaltan para que nos comportemos de una manera apropiada y eficiente y no ser engañados. Tal como lo vimos en los capítulos 1 y 2, hay tanta información que debemos tener cuidado de mantener un sentido de proporción y equilibrio en la forma en que la interpretamos y reaccionamos a ella. Un sentido de proporción acerca de la realidad puede medirse con números o imágenes.

Para transmitir un sentido de proporción, describimos numéricamente la distribución de las puntuaciones de una variable con frecuencias porcentuales, como lo hicimos en el capítulo 2. Las distribuciones numéricas, sin embargo, tienen sentido sólo si una persona tiende a pensar de manera proporcional. Los gráficos constituyen un excelente soporte a la célebre máxima:

“Una imagen dice más que mil palabras.” Cuando algunas de esas palabras son numéricas, una imagen hace más claros cientos de palabras y cálculos.

Debido a lo fáciles que son actualmente los programas de cómputo, los medios de comunicación masiva nos bombardean con gráficos, gráficos de pastel y pictográficos (fotografías de objetos, iconos o mapas sombreados). A veces, los programas de cómputo toman vida propia y crean un gráfico de forma incorrecta. En términos del control del error, es importante conocer los cálculos matemáticos que hay detrás de la construcción gráfica y no sólo confiar en los programadores.

Los diseños gráficos y pictóricos se eligen con base en (1) el nivel de medición de una variable, (2) los objetivos y los aspectos relevantes del estudio y (3) el público a quien se dirigen. Para las audiencias públicas, los gráficos sencillos y a todo color funcionan mejor y brindan una perspectiva global de los estadísticos descriptivos, tales como porcentajes y promedios. En contraste, los públicos compuestos por especialistas están acostumbrados a los estadísticos inferenciales, aquéllos diseñados para explicar y probar hipótesis. Junto con las tablas estadísticas, los gráficos nos ayudan a discernir las formas de las distribuciones de frecuencias. Incluso los gráficos descriptivos alertan a un analista sobre fuentes de error potenciales que puedan influir en el análisis realizado.

Lineamientos para graficar

Las presentaciones gráficas deben cumplir con algunas reglas y lineamientos simples, los cuales también se aplican a las tablas y a la elaboración de reportes.

Lineamientos para graficar

1. Elige el diseño con base en a) el nivel de medición de una variable, b) los objetivos del estudio y c) el público a quien se dirige.
2. Ante todo, una buena presentación gráfica tiene que ser clara y entendible. Debe simplificar, no complicar.
3. Un gráfico o diagrama requiere explicarse por sí mismo y transmitir información, sin hacer referencia a un texto o a alguien que lo explique. La selección cuidadosa de títulos, descripción de la escala, subtítulos y otras leyendas contribuyen a lograr este objetivo. Somete cada gráfico o tabla a la prueba de “perdido en el estacionamiento”. Pregúntate: si este gráfico fuera abandonado en un estacionamiento, ¿podría tomarlo un perfecto extraño e interpretarlo?
4. Antes de decidirte sobre el tipo de presentación pictórica (por ejemplo, gráfico de pastel contra gráfico de barras), elabora bosquejos con varias opciones. Los programas de cómputo hacen esto en forma relativamente fácil. Para ampliar las alternativas, solicita opiniones y consulta otros materiales, tales como informes organizacionales.
5. Adhiérete a los principios de inclusividad (capítulo 2). Anota al pie de página cualquier excepción.
6. Si los datos no son tuyos, indica la fuente de los mismos al final de la tabla.

Graficación de datos nominales/ordinales

Gráficos de pastel

Un estilo sencillo de presentación para datos nominales/ordinales es el gráfico de pastel. Un **gráfico de pastel** es un círculo que se divide (o rebanan) desde su punto central, donde cada rebanada representa la frecuencia proporcional de determinada categoría. Todos hemos rebanado pasteles y, a veces, no obtenemos una porción justa. Cuando un investigador quiere ofrecer un sentido de proporción respecto de una variable nominal/ordinal, los gráficos de pastel son especialmente útiles, pues representan de manera equitativa el tamaño relativo o desigualdad entre las categorías. El tamaño relativo de una rebanada de pastel es una forma de pensamiento proporcional, con la cual todos estamos familiarizados.

Gráfico de pastel Círculo que se divide (o rebanan) desde su punto central, donde cada rebanada representa la frecuencia proporcional de determinada categoría de una variable nominal/ordinal. Es especialmente útil para transmitir un sentido de equidad, tamaño relativo o desigualdad entre las categorías.

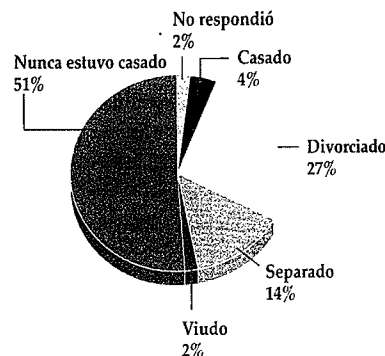
La figura 3-1 muestra la distribución del estatus marital en una muestra de 161 personas sin hogar. El área dentro del círculo entero representa el 100 por ciento de los sujetos en la muestra. El área de una rebanada indica el porcentaje en una categoría específica. Es fácil percibir que más de la mitad de los encuestados nunca estuvieron casados y que una porción sustancial se había divorciado. La revelación más sorprendente es la parte pequeña correspondiente a “casado”.

Mientras un programa de cómputo sin dificultad produjo el gráfico de pastel de la figura 3-1, para establecer la relación entre el gráfico de pastel y la distribución de frecuencias del estado civil elaboremos el gráfico a mano. El primer paso para elaborar cualquier gráfico consiste en determinar la distribución de frecuencias de la variable. Del mismo modo, con los gráficos de pastel calculamos la frecuencia proporcional y la frecuencia porcentual de cada categoría. Las frecuencias proporcionales, junto con el conocimiento sobre las dimensiones de un círculo, sirven para calcular el tamaño de las rebanadas.

La división correcta del pastel depende de saber que los ángulos que cortan un círculo desde su centro se miden en grados con un transportador (regla circular con forma de media luna). Sin importar el tamaño de un círculo, su circunferencia se define por tener una distan-

FIGURA 3-1

Diagrama del estatus marital de personas sin hogar, $n = 161$



cia circular total de 360 grados ($^{\circ}$). Medio círculo tiene 180° o una proporción de 0.5 por 360 grados; y un cuarto de círculo, 90° o $\frac{1}{4}$ de 360 grados. Estos puntos de referencia del círculo se ilustran en la figura 3-2. Cualquier parte del pastel se rebana multiplicando la proporción en una categoría por 360 grados.

La tabla 3-1 muestra los cálculos para el gráfico de pastel que aparece en la figura 3-1. ¿Cómo se le asigna a cada categoría su porción del pastel? La porción de una categoría es la frecuencia proporcional p por 360° , la circunferencia total del pastel. Si 0.51 (51 por ciento) de los encuestados nunca habían sido casados, deben “obtener” 51 por ciento del pastel, es decir un valor de 185° , o simplemente más de la mitad. Después que los 360 grados hayan sido asignados a todas las categorías, se utiliza un transportador para trazar en el papel las porciones correctamente proporcionadas. Para lograr mayor claridad, se anotan los porcentajes en el gráfico de pastel. Si el gráfico se presenta ante una audiencia pública, redondeamos a un porcentaje entero (es decir, la posición de las unidades).

Interpretación de los gráficos de pastel Regresemos a la figura 3-1, que presenta la distribución de estado civil para una muestra de 161 personas sin hogar, y sistemáticamente interpretemos su significado. Primero nos concentramos en las rebanadas más grandes, que representan categorías que ocurren con las mayores frecuencias. En la figura 3-1 es fácil ver que más de la mitad de quienes respondieron y no tienen casa nunca habían sido casados y muchos de ellos eran divorciados o separados. Segundo, comparemos entre sí los tamaños de las rebanadas. Un porcentaje de personas sin casa que nunca habían sido casados era más alto que todas las otras categorías combinadas. Las personas que no tenían pareja, es decir que nunca se habían casado, divorciado o separado, constituían un segmento grande de personas sin hogar. Tercero, comparemos los resultados con otras poblaciones. Busquemos rebanadas inesperadamente pequeñas o grandes de esta población en comparación con otras poblaciones. Lo más sorprendente de la figura 3-1 es la pequeña pieza correspondiente a “casados”. Por ejemplo, en la población general de adultos, 56.7 por ciento de los adultos de más de 18 años son casados en comparación con sólo el 4 por ciento de esta muestra de personas sin hogar (U.S. Bureau of the Census 2003). En resumen, no tener casa no es benéfico para el estado civil.

En la tabla 3-1 nótese que el porcentaje de la categoría “Nunca estuvo casado” se redondea hacia abajo a 51 por ciento en lugar de hacia arriba a 52 por ciento. Es más, asignamos a esta categoría 185° en lugar de 186° . Estos ajustes por error de redondeo son necesarios para impedir que el total de grados del círculo exceda los 360°, pues un círculo tiene un espacio definido de esa magnitud. Ajustar la categoría “Nunca estuvo casado” tiene el efecto más pequeño de error en comparación con el ajuste en otras categorías.

FIGURA 3-2

Grados de un círculo: un cuarto, medio y círculo completo.

Puntos de referencia

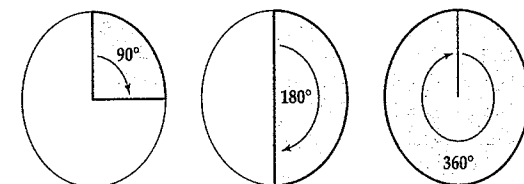


TABLA 3-1 Hoja de cálculo para construir un gráfico de pastel: distribución de estado civil para una muestra de 161 personas sin hogar

Especificaciones		Cálculos		
Estado civil	f	p	(p)(360°)	Porcentaje (%)
Nunca estuvo casado	83	.5155	185°	51
Divorciado	43	.2671	96	27
Separado	22	.1366	49	14
Casado	7	.0435	16	4
Viudo	3	.0186	7	2
No respondió	3	.0186	7	2
Totales	161	.9999*	360°	100

*El total no sumó 1.0000 por el error de redondeo.

Los paquetes de software (por ejemplo el *SPSS for Windows*, que es una opción con este texto) ofrecen una amplia gama de estilos para gráficos de pastel. Una o más rebanadas pueden presentarse en relieve o "piezas desarmadas", y se pueden presentar pares de gráficas de pastel para comparar grupos o periodos.

Forma de construir e interpretar una gráfica de pastel

Para construir una gráfica de pastel:

1. Elabora una tabla de distribución de frecuencia con los siguientes encabezados:

Categoría	f	p	(p)(360°)	(%)
-----------	---	---	-----------	-----

donde

"Categoría" = nombre de la categoría de una variable nominal/ordinal,

f = frecuencia de casos (o número de casos) de una categoría,

p = p [del n total en una categoría] = (f de categoría)/n, con n = tamaño muestral,

(p)(360°) = grados para cada rebanada,

% = porcentaje [del n total en una categoría] = (p)(100).

2. Traza un círculo y pon un punto en su centro. Traza una recta del punto al círculo. Pon un transportador sobre esta recta, marca el número de grados para la primera categoría y traza una recta para crear la rebanada del pastel. Pon un transportador sobre esta segunda recta, marca el número de grados para la segunda categoría, y así sucesivamente. Asegúrate de que haya exactamente 360°.

3. Marca con toda claridad cada rebanada e indica el porcentaje de casos que representa. Asegúrate de que las leyendas sean horizontales (es decir, no las ajustes a la forma circular de la rebanada). Utiliza tu juicio de dibujante para poner las leyendas dentro o fuera de las rebanadas.
4. Titula adecuadamente el gráfico de pastel. Identifica la fuente de datos en la parte inferior del gráfico.

Para interpretar un gráfico de pastel:

1. Concéntrate en las rebanadas más grandes del pastel. Estipula las categorías que se presentaron con mayor frecuencia.
2. Compara entre sí los tamaños de rebanadas.
3. Si es apropiado, compara los resultados con otras poblaciones. Busca rebanadas inesperadamente pequeñas o grandes (por ejemplo, el pequeño porcentaje de personas casadas entre adultos sin hogar de la figura 3-1).

Gráficos de barras

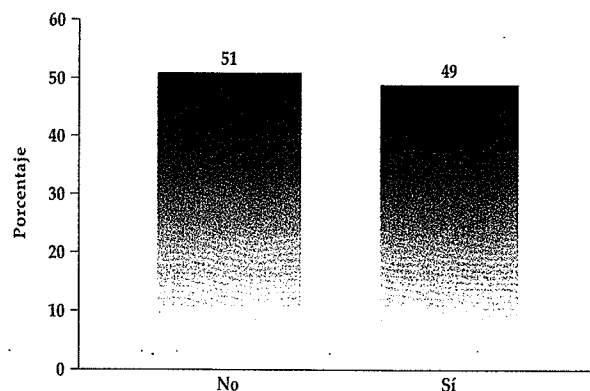
Otra manera de graficar datos nominales/ordinales consiste en utilizar un gráfico de barras. Un **gráfico de barras** se compone de una serie de barras verticales u horizontales, donde la longitud de la barra representa la frecuencia porcentual de una categoría de una variable nominal/ordinal. Al igual que una rebanada de un gráfico de pastel, el área de una barra determinada por su longitud transmite un sentido de frecuencia proporcional de una categoría. Los gráficos de barras se construyen sobre dos ejes: uno trazado horizontalmente (el de las *abscisas*) y el otro colocado en forma vertical (el de las *ordenadas*). En otras palabras, las dos líneas se unen en un ángulo de 90° o ángulo recto. Las categorías de una variable se sitúan en un eje y las marcas para los porcentajes en el otro. Sólo necesitamos calcular la frecuencia porcentual de cada categoría para elaborar un gráfico de barra.

Gráfico de barras Serie de barras verticales u horizontales, donde la longitud de la barra representa la frecuencia porcentual de una categoría de una variable nominal/ordinal. Los gráficos de barras son especialmente eficaces para ilustrar una competencia entre categorías.

Interpretación de gráficos de barras La figura 3-3 presenta un gráfico de barras de la distribución de frecuencia porcentual de empleo de adultos sin hogar en la semana anterior a una entrevista. Primero, observa la altura de las barras. La barra más alta es la categoría con la frecuencia más alta. Haz algún comentario sobre el orden de las categorías. En la figura 3-3 vemos que más personas sin hogar no trabajaron en la semana previa que las que sí trabajaron. Segundo, compara las barras y comenta sobre algunas de estas que sean especialmente altas o cortas. En la figura 3-3 notamos que las barras tienen más o menos la misma altura. Tercero, si es apropiado, compara los resultados con otras poblaciones. Busca barras inesperadamente altas o cortas. Un investigador experto señalaría que los resultados contradicen

FIGURA 3-3

Gráfico de barras de adultos sin hogar en la semana anterior a una entrevista, $n = 161$



la idea comúnmente admitida de que las personas sin hogar son vagos o indolentes. Si se consideran las circunstancias de quienes no tienen hogar, uno esperaría un porcentaje mucho más bajo de trabajo reciente que el que revela o se deja ver en esta muestra. A este respecto, entonces, el porcentaje de quienes trabajan es más alto de lo esperado. En resumen, la similitud de las alturas de las barras rápidamente da a entender que hay casi tantos sin hogar que trabajan como los que no trabajan.

Forma de construir e interpretar un gráfico de barras

Para construir un gráfico de barras:

1. Elabora una tabla de distribución de frecuencias con los encabezados siguientes:

Categoría	f	p	(%)
-----------	-----	-----	-----

donde

“Categoría” = nombre de la categoría de una variable nominal/ordinal,

f = frecuencia de casos (o número de casos) en una categoría,

$p = p[\text{del } n \text{ total de una categoría}] = (f \text{ de categoría})/n$, con n = tamaño muestral,

% = porcentaje [del n total de una categoría] = $(p)(100)$.

2. Traza el eje horizontal del gráfico de barras, con un ancho apropiado para el número de barras.
3. Traza el eje vertical. Observa la frecuencia más alta (f) o frecuencia porcentual de la tabla de distribución de frecuencia y escribe marcas en el eje que vayan desde cero hasta un poco más de la frecuencia más alta.

4. Traza las barras con un ancho que sea visualmente atractivo. Puedes escoger el orden de las barras de la más alta a la más baja. Separa las barras de cada categoría. Utiliza frecuencias (f) o frecuencias porcentuales sobre el eje vertical como marcadores de altura de barras. Bajo las barras incluye una leyenda clara con los nombres de las categorías.
5. Escribe un título apropiado para gráfico de barras. Asegúrate de que los títulos que apliques a los ejes sean precisos y claros. En la parte inferior del gráfico identifica la fuente de los datos.

Para interpretar un gráfico de barras:

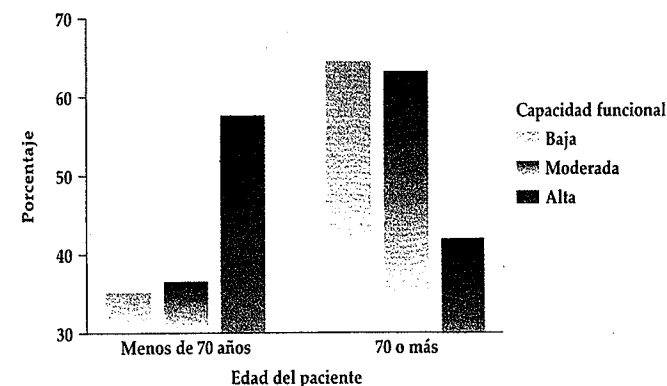
1. Observa la altura de las barras. La barra más alta indica la categoría que tiene la frecuencia más alta. Haz comentarios sobre el orden de las categorías.
2. Compara las barras y comenta sobre cualquiera que sea especialmente alta o corta.
3. Si es apropiado, compara los resultados contra otras poblaciones. Busca barras especialmente altas o cortas.

La figura 3-4 presenta un gráfico de barras “agrupadas”, que es muy útil para comparar dos o más grupos en una variable nominal/ordinal. Esta figura compara la habilidad funcional de 104 pacientes de un hospital de veteranos y deja ver cómo la baja capacidad funcional es tan característica en veteranos enfermos mayores de 70 años.

Por último, el eje vertical de un gráfico de barras no siempre mide simplemente números o porcentajes. Entre las categorías de una variable nominal/ordinal los gráficos de barras se pueden usar para expresar cantidades relativas de cualquier variable. Por ejemplo, el ejercicio 3B-2, que aparece en los ejercicios al final de este capítulo, se refiere a un gráfico de barras sobre el consumo de alcohol para cinco países europeos. La variable nominal es el país, y los nombres de países se aplican por todo el eje horizontal del gráfico de barras. En el eje vertical, en lugar de indicar un número o porcentaje, la escala será “litros de alcohol consumido”. Los valores a graficar se ven directamente en la tabla de ese ejercicio. Del mismo modo, podríamos construir un gráfico de barras que haga una comparación de los ingresos medios de estos cinco países con las cantidades en dólares indicadas en el eje vertical.

FIGURA 3-4

Gráfico de barras agrupadas referente a la capacidad funcional de pacientes de un hospital de veteranos por edad, $n = 104$



Grificación de variables de intervalo/razón

Histogramas

Un histograma es un tipo de gráfico que se utiliza con variables de intervalo/razón. Una de estas variables de razón, que definimos como X , es la de evaluaciones de rendimiento de combustible proporcionadas por la Environmental Protection Agency (U.S. Department of Energy, 2004). Estas evaluaciones se estiman en millas por galón (MPG), que se fijan en modelos nuevos de vehículos. De nueva cuenta, el primer paso para cualquier gráfico es elaborar una distribución de frecuencias. La hoja de trabajo de cálculo de la tabla 3-2 presenta la distribución de frecuencia de evaluaciones de rendimiento de combustible, para conducción en la ciudad de modelos de autos compactos de cuatro cilindros del año 2004 (excluyendo modelos híbridos de gasolina/eléctricos). Nuestro interés está en cómo se agrupan las puntuaciones y en cómo se dispersan. Con toda facilidad podemos ver, por ejemplo, que la evaluación mínima fue 18 MPG, y la máxima, de 38 MPG. Si observamos las evaluaciones de rendimiento de combustible con alta frecuencia (es decir, aquellas para las que f es grande),

TABLA 3-2 Hoja de cálculo para construir histogramas y polígonos: distribución de frecuencia de evaluaciones de rendimiento de combustible para conducción en ciudades, en millas por galón (MPG); 106 modelos de autos compactos; modelo 2004

Especificaciones		Cálculos
X		
Evaluación de rendimiento de combustible (MPG)	f	Límites reales
18	1	17.5-18.5
20	4	19.5-20.5
21	6	20.5-21.5
22	19	21.5-22.5
23	10	22.5-23.5
24	17	23.5-24.5
25	9	24.5-25.5
26	13	25.5-26.5
27	5	26.5-27.5
28	5	27.5-28.5
29	4	28.5-29.5
30	1	29.5-30.5
31	1	30.5-31.5
32	3	31.5-32.5
33	3	32.5-33.5
35	2	34.5-35.5
36	1	35.5-36.5
38	2	37.5-38.5
Total	106	

Fuente: U.S. Department of Energy, 2004.

podemos ver que muchos modelos de autos compactos están proyectados para rendir entre 22 y 26 MPG en condiciones de conducción en ciudades. (Para personas acostumbradas a medir el rendimiento de combustible en unidades métricas, multiplica MPG por 0.42 para obtener el equivalente en kilómetros por litro de gasolina.)

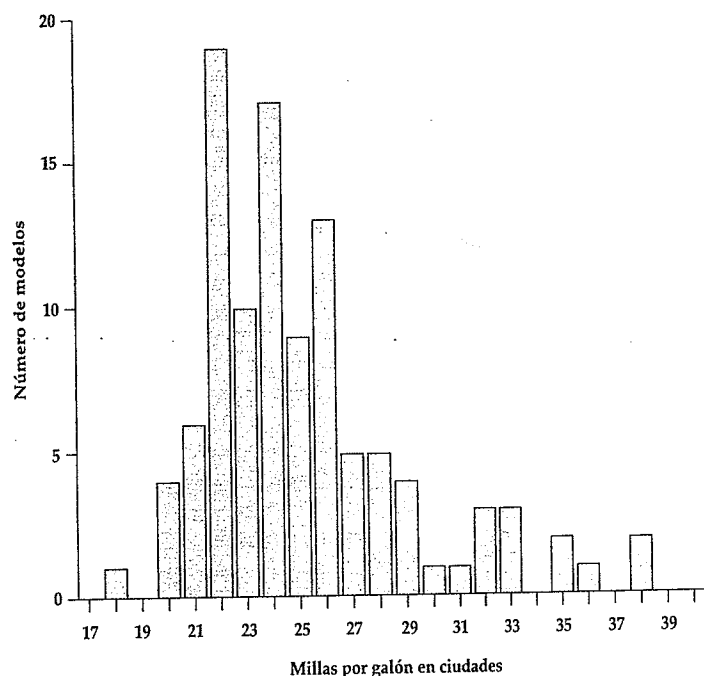
Al hacer un gráfico de datos, obtenemos un sentido de proporción incluso mejor acerca de cómo las evaluaciones están distribuidas para autos compactos. La figura 3-5 presenta las evaluaciones de rendimiento de combustible de la tabla 3-2 en forma de un histograma de frecuencia. Un **histograma de frecuencias** es un gráfico de 90° que presenta las puntuaciones de una variable a lo largo del eje horizontal, y la frecuencia de cada puntuación en una columna paralela al eje vertical. En otras palabras, se grafica X sobre el eje horizontal y f en el vertical. Un histograma es semejante a un gráfico de barras, excepto que las columnas de un histograma se tocan entre sí (a menos que una puntuación tenga una frecuencia de cero casos, en cuya circunstancia habrá una columna faltante). Los puntos graficados sobre el eje horizontal reciben el nombre de marcas y representan las puntuaciones X (en este ejemplo, evaluaciones en MPG). Nótese que las marcas más bajas y más altas, 17 y 39 MPG, respectivamente, se prolongan un poco por debajo de las puntuaciones mínima y máxima observadas de 18 y 38 MPG. Para trazar y aplicar una leyenda al eje vertical, en la tabla de distribución de frecuencia identifica la frecuencia más alta (f) registrada. En la tabla 3-2 ésta es una frecuencia de 19 para una puntuación de 22 MPG. Aplica una leyenda al eje vertical de 0 a un poco más de 19. El ancho de cada columna del histograma es igual. Las columnas que se tocan entre sí son los límites reales de cada puntuación. Por ejemplo, las puntuaciones de la tabla 3-2 están redondeadas al entero más cercano; en otras palabras, cada modelo evaluado a 22 MPG no rindió exactamente 22 MPG en los cálculos de la Environmental Protection Agency. Tomar en cuenta los límites reales ensancha las columnas en forma tal que las columnas adyacentes se tocan entre sí y satisfacen el principio de inclusividad.

Histograma de frecuencia Es un diagrama de 90 grados que presenta las puntuaciones de una variable de intervalo/razón a lo largo del eje horizontal, y la frecuencia de cada puntuación en una columna paralela al eje vertical.

Interpretación de histogramas ¿Qué mensaje expresa el histograma de la figura 3-5? Hay varias características de los histogramas que transmiten información. Primero, observa la altura de las columnas. Podemos ver que la evaluación de rendimiento de combustible que se presenta con más frecuencia es 22 MPG, y las evaluaciones de 24 y 26 millas también son bastante comunes. Segundo, busca grupos de puntuaciones y ve si hay una "tendencia central", es decir, un valor de puntuación alrededor del que se centra la distribución. Podemos ver que el grueso de los modelos rinden entre 20 y 29 MPG. Además, con excepción de unas pocas puntuaciones especialmente altas, las evaluaciones tienden a centrarse alrededor de 24 MPG. Tercero, busquemos la simetría o equilibrio en la distribución de puntuaciones. Podemos ver que este histograma no es simétrico. El grueso de las puntuaciones X se apoya sobre el extremo inferior de la distribución, y unos cuantos modelos, los que rinden arriba de 30 MPG, tienen evaluaciones de rendimiento de combustible especialmente altas. (En el capítulo 4 definiremos la forma de esta distribución de puntuación como un sesgo a la derecha.)

FIGURA 3-5

Histograma de frecuencias de las evaluaciones de rendimiento de combustible de 106 modelos de autos compactos; modelo 2004



Fuente: U.S. Department of Energy, 2004.

Forma de construir e interpretar histogramas

Para construir un histograma:

1. Elabora una tabla de distribución de frecuencias con los encabezados siguientes:

Puntuación (X)	f	Límites reales de puntuación
----------------	---	------------------------------

donde

Puntuación = puntuación de una variable de intervalo/razón

f = frecuencia de casos (o número de casos) en una puntuación

2. Calcula los límites reales de cada puntuación. (En el capítulo 2 puedes revisar los límites reales.)
3. Traza un eje horizontal del histograma. Observa las puntuaciones más baja y más alta en la tabla de distribución de frecuencias. Escribe marcas sobre el eje y aplica de conformidad los valores de X ; deja espacio adicional en cada extremo del eje fuera de los valores de las puntuaciones más baja y más alta.

4. Traza el eje vertical. Observa la frecuencia más alta (f) en la tabla de distribución de frecuencias y escribe marcas en el eje que vayan desde cero hasta un poco más de la frecuencia más alta.
5. Traza las columnas usando límites reales como marcadores para anchos de columna y frecuencias (f) como marcadores para alturas de columna.
6. Aplica un título preciso al histograma. Asegúrate de que las leyendas de ejes sean correctas y claras. Identifica la fuente de datos en la parte inferior del gráfico.

Para interpretar un histograma:

1. Observa la altura de las barras. La columna más alta indica el valor de la puntuación de X que tenga la frecuencia más alta (f).
2. Busca grupos de puntuaciones y ve si hay una "tendencia central", un valor de puntuación de X alrededor del que se centra la distribución.
3. Busca la simetría o equilibrio en la distribución de las puntuaciones. ¿Las puntuaciones tienden a ubicarse en forma homogénea alrededor de una puntuación central, o son puntuaciones especialmente bajas o altas (como en la figura 3-5)?

Polígonos y gráficos de líneas

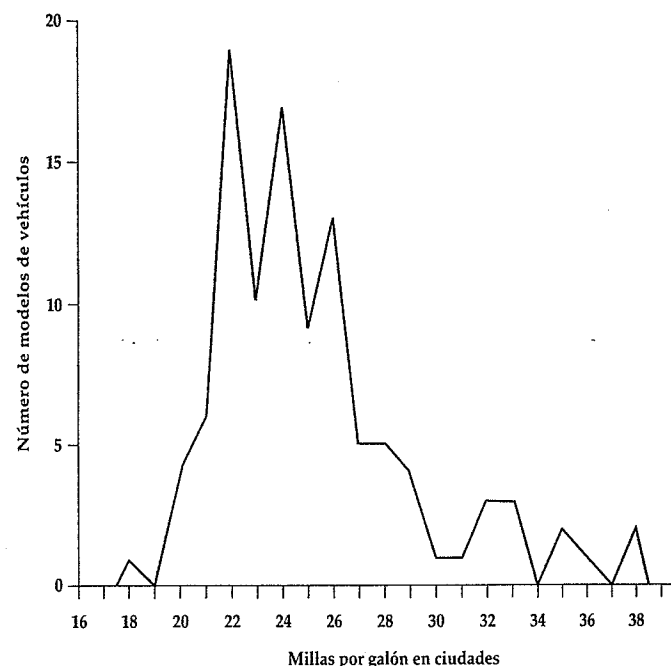
Otra técnica gráfica para representar variables de intervalo/razón es el polígono de frecuencia o gráfico de líneas. Un **polígono de frecuencias** es un diagrama de 90 grados con puntuaciones de intervalo/razón señaladas en el eje horizontal o línea base, y las frecuencias de las puntuaciones están representadas por las alturas de puntos localizados sobre las puntuaciones y enlazados mediante líneas rectas. Los ejes de un polígono se diseñan como los de un histograma. Los valores de X están marcados en el eje horizontal o línea base. Las frecuencias (f) se grafican en el eje vertical, pero, para indicar la frecuencia de la variable en una puntuación particular, usamos puntos en lugar de columnas y enlazamos los puntos para obtener una línea gráfica. La figura 3-6 es el polígono para las evaluaciones de rendimiento de combustible de la tabla 3-2, que usamos para el histograma de la figura 3-5. Mientras que los histogramas atraen la atención a las columnas más altas, donde se encuentra el grueso de las puntuaciones, los polígonos comunican un sentido de tendencia o movimiento. Esto es, observamos el flujo de picos y valles en la línea gráfica cuando comparamos las evaluaciones de rendimiento de combustible de la más baja a la más alta por toda la línea de base.

Polígono de frecuencias (gráfico de líneas) Gráfico de 90 grados con la puntuación de intervalo/razón trazada sobre el eje horizontal, y las frecuencias de puntuación descritas por las alturas de puntos localizados sobre puntuaciones y enlazados por líneas rectas.

Interpretación de los polígonos (gráficos de líneas) ¿Qué mensaje expresa el polígono de la figura 3-6? Las notables características de los polígonos son semejantes a las de los histogramas. Primero, identifica el pico más alto del polígono. Este pico se manifiesta para una evaluación de rendimiento de combustible de 22 MPG y ésta es la puntuación que

FIGURA 3-6

Polígono de frecuencia de la distribución de evaluación de combustible en la ciudad en millas por galón (MPG) para autos compactos, modelo 2004



Fuente: U.S. Department of Energy, 2004.

ocurre con más frecuencia. Los picos ligeramente más bajos se muestran para evaluaciones de rendimiento de combustible de 24 y 26 MPG, revelando estos valores que también se presentan con frecuencia. Segundo, busca una extensión de espacio bajo la línea gráfica y ve si hay una tendencia central. Podemos ver que casi toda el área bajo la línea gráfica se apoya entre las puntuaciones de 20 y 29 MPG, y las evaluaciones tienden a centrarse alrededor de 24 MPG. Hay unas pocas puntuaciones especialmente altas, con más de una puntuación de 30 MPG. Tercero, busca la simetría o equilibrio en la distribución de las puntuaciones y busca una tendencia en la forma de la línea gráfica. Ve si las colas de la línea gráfica se extienden de manera uniforme. En la figura 3-6 la “cola de dragón” que se prolonga a la derecha hace resaltar el desequilibrio de la distribución. En tanto que casi todos los modelos de autos compactos tienen evaluaciones de rendimiento debajo de 30 MPG, unos cuantos modelos son especialmente eficientes. En general, el polígono expresa el mensaje de que conforme nos movemos hacia arriba en la escala de evaluaciones de rendimiento de combustible, hacia un mayor rango de MPG, existen menos modelos.

Los polígonos de frecuencia son especialmente útiles para comparar dos o más muestras. Por ejemplo, comparemos las distribuciones de evaluaciones de rendimiento de combustible para autos compactos contra los vehículos utilitarios de tracción en las cuatro ruedas (SUV). La tabla 3-3 da las distribuciones de frecuencia y frecuencia porcentual para ambos tipos de vehículos. Nótese que difieren los tamaños muestrales de los tipos de vehículos. Hay 106 modelos de autos compactos pero sólo 68 modelos de los SUV. Si usamos frecuencias sin elaborar para construir los polígonos, el polígono para los autos compactos más numerosos hará empuqueñecer el polígono para los SUV. Entonces, usamos un común denominador de

Forma de construir e interpretar polígonos (gráficos de línea)

Para construir un polígono:

1. Elabora una tabla de distribución de frecuencias con los siguientes encabezados (exactamente como la tabla empleada para construir histogramas):

Puntuación (X)	f	Límites reales de puntuación
----------------	---	------------------------------

donde

Puntuación = puntuación de una variable de intervalo/razón

f = frecuencia de casos (o número de casos) para una puntuación

2. Calcula los límites reales de cada puntuación X. (En el capítulo 2 puedes consultar sobre los límites reales.)
3. Traza un eje horizontal o “línea de base” del polígono. Observa las puntuaciones más baja y más alta en la tabla de distribución de frecuencia. Escribe marcas sobre el eje y aplica de conformidad los valores de X; deja espacio adicional en cada extremo del eje fuera de los valores de las puntuaciones más baja y más alta.
4. Traza el eje vertical. Observa la frecuencia más alta (f) en la tabla de distribución de frecuencia y escribe marcas en el eje que vayan desde cero hasta un poco más de la frecuencia más alta.
5. A partir del valor más bajo de X y avanzando hasta el más alto, traza puntos arriba de cada valor de X hasta la altura de su frecuencia (f).
6. Une los puntos con líneas rectas. *Atención:* si un valor de X tiene una frecuencia de cero (por ejemplo, la calificación de 34 MPG en la tabla 3-2), la línea se prolonga por debajo de la línea de base.
7. Cierra los extremos de la línea del gráfico. Con una línea entre su punto y su límite real más bajo, une el valor más bajo de X a la línea de base. Con una línea entre su punto y su límite real superior, une el valor más alto de X a la línea de base.
8. Asigna un título apropiado al polígono. Asegúrate de que las leyendas de los ejes sean correctas y claras. Identifica la fuente de los datos en la parte inferior del gráfico.

Para interpretar un polígono:

1. Busca picos. El pico más alto indica el valor de X con la frecuencia más alta.
2. Busca una extensión de espacio bajo la línea gráfica para ver si hay agrupaciones de puntuaciones y para ver si hay una tendencia central.
3. Busca simetría o equilibrio en la distribución de puntuaciones. Busca una tendencia en la forma de la línea del gráfico. Ve si las colas del gráfico están situadas de manera homogénea alrededor de una puntuación central. Si no es así, observa puntuaciones especialmente bajas o altas al identificar cuál cola se prolonga.

100 al calcular las frecuencias porcentuales. La figura 3-7 presenta los dos polígonos con “Porcentaje de modelos de vehículos” estipulado en el eje vertical. Una vez que estas dos distribuciones sean descritas gráficamente juntas, sus diferencias quedan bien claras. Los picos de las dos líneas expresan una diferencia en tendencia central o promedio, que es el tema de nuestro siguiente capítulo. Los vehículos SUV tienden a evaluaciones de rendimiento relativamente bajas, en una franja más bien angosta que va de 15 a 29 MPG. En contraste, los autos compactos tienen rendimientos más bien altos y una franja más ancha de 20 a 29 MPG. Una de las características más sorprendentes de este gráfico es el pequeño traslape entre las dos líneas. Los SUV más eficientes tienen evaluaciones de rendimiento iguales a los autos compactos menos eficientes.

TABLA 3-3 | Comparación de las evaluaciones de rendimiento de combustible para operación en ciudades, en millas por galón (MPG), para modelos de autos compactos y vehículos utilitarios con tracción en las 4 ruedas (SUV); modelo 2004.

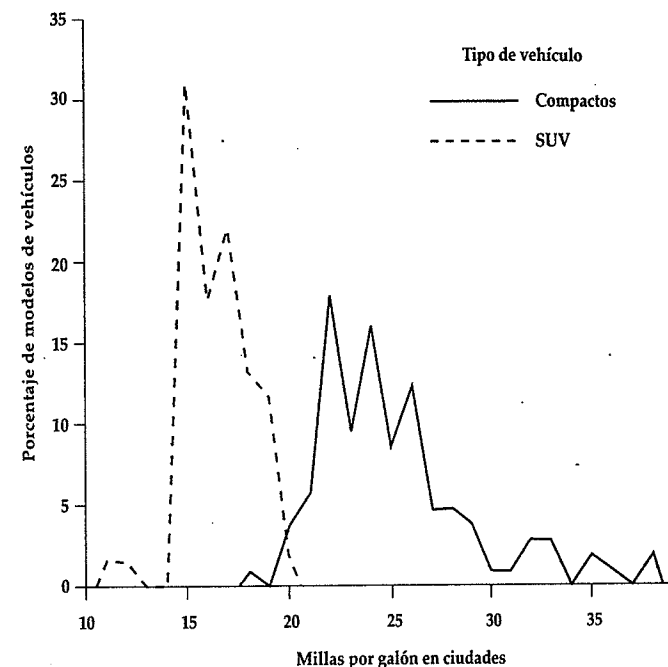
Rendimiento de combustible (MPG)	Modelos de autos compactos		Modelos SUV	
	f	Porcentaje (%)	f	Porcentaje (%)
11	0	0.0	1	1.5
12	0	0.0	1	1.5
15	0	0.0	21	30.9
16	0	0.0	12	17.6
17	0	0.0	15	22.1
18	1	0.9	9	13.2
19	0	0.0	8	11.8
20	4	3.8	1	1.5
21	6	5.7	0	0.0
22	19	17.9	0	0.0
23	10	9.4	0	0.0
24	17	16.0	0	0.0
25	9	8.5	0	0.0
26	13	12.3	0	0.0
27	5	4.7	0	0.0
28	5	4.7	0	0.0
29	4	3.8	0	0.0
30	1	0.9	0	0.0
31	1	0.9	0	0.0
32	3	2.8	0	0.0
33	3	2.8	0	0.0
35	2	1.9	0	0.0
36	1	0.9	0	0.0
38	2	1.9	0	0.0
Totales	106	99.8	68	100.1*

*Los porcentajes totales no suman 100 por error de redondeo.

Fuente: U.S. Department of Energy, 2004.

FIGURA 3-7

Comparación de las evaluaciones de rendimiento de combustible para conducción en ciudades, en millas por galón, para 106 modelos de autos compactos y 68 vehículos SUV; modelo 2004



Fuente: U.S. Department of Energy, 2004.

Por último, histogramas y polígonos son útiles para identificar puntuaciones inusualmente bajas o altas en una distribución. Por ejemplo, en la figura 3-7 observa que hay dos modelos SUV con rendimientos de combustible especialmente bajos (11 y 12 MPG). Del mismo modo, hay unos pocos modelos de autos compactos con rendimientos especialmente altos, arriba de 34 MPG. Estas puntuaciones poco comunes reciben el nombre de *valores extremos*. Además de atraer nuestra atención a puntuaciones “extremosas”, es importante identificar resultados extremos y considerar sus efectos sobre el análisis estadístico inferencial.

Uso de gráficos en la estadística inferencial y su aplicación en la investigación

Los gráficos sirven principalmente para propósitos descriptivos en audiencias públicas. En la investigación científica y en la estadística inferencial, no obstante, a veces graficamos una variable para familiarizarnos con su distribución y prepararla para un análisis posterior. Para agilizar esta fase de preparación, por medio de una computadora generamos histogramas y polígonos (gráficos de líneas). En estadística inferencial, tales gráficos son muy útiles para detectar puntuaciones atípicas en una distribución. Por ejemplo, un gráfico que ilustra las tasas de divorcio en 50 estados de la Unión Americana revela que Nevada, con sus leyes que favorecen el divorcio, es notablemente diferente. Un caso atípico se llama **puntuación desviada o valor extremo**, que es una puntuación notablemente diferente de las otras en la distribución de puntuaciones. Como lo veremos más adelante, los valores extremos distorsionan los cálculos estadísticos, como —por ejemplo— los promedios. Si las distorsiones son grandes, es necesario desechar o ajustar estas puntuaciones desviadas. Por ejemplo, en el

caso de los porcentajes de divorcio el procedimiento adecuado podría ser desechar los datos de Nevada. Además, informamos al lector que se trata de un estado excepcional, digno de un estudio de caso (análisis individualizado), y manifestamos que nuestras conclusiones sólo se aplican a los restantes 49 estados. Otra manera de modificar los efectos distorsionantes de los valores extremos consiste en ajustar matemáticamente las puntuaciones extremas. Un método requiere tomar el logaritmo de las puntuaciones, una transformación matemática que comprime las puntuaciones en un rango menor. Un segundo método radica simplemente en reducir el valor extremo a la siguiente puntuación más baja o más alta, que es un procedimiento llamado "truncamiento". Una recomendación final: es impropio omitir o ajustar valores extremos simplemente porque no se ajustan al patrón esperado. El investigador debe explicar con toda claridad las razones teóricas y prácticas para realizar tales ajustes. En próximos capítulos se abundará sobre los valores extremos.

Puntuación desviada o valor extremo Puntuación que es notablemente diferente de las otras en la distribución de puntuaciones.

Insensatez y falacias estadísticas: distorsión gráfica

Los gráficos y los diagramas ofrecen mapas mentales de conjuntos grandes de datos. Los procedimientos para diseñar gráficos son normativos, es decir, diferentes personas tienen ideas distintas respecto de aquello que agrada al ojo. En otras palabras, la presentación gráfica es casi un arte.

Puesto que las normas de presentación de datos son tanto estéticas como técnicas, a menudo son poco claras. Por ejemplo, ¿qué tan amplias deben ser las barras en un gráfico? ¿Debemos utilizar varios colores en las barras? Las convenciones que se aplican a éstas y otras cuestiones similares son flexibles y con frecuencia siguen una moda. El arte inspira la creatividad y la individualidad.

No obstante, cuando las reglas son poco claras es fácil transgredirlas, intencional o involuntariamente. Por ejemplo, con la amplia variedad de programas gráficos de computadora que existen en la actualidad, los usuarios a menudo están dispuestos a dejar los detalles del

FIGURA 3-8

Cartel presentado a Mortimer Mainstreet por su personal de campaña, para informar sobre los resultados de la más reciente encuesta en la campaña para gobernador.

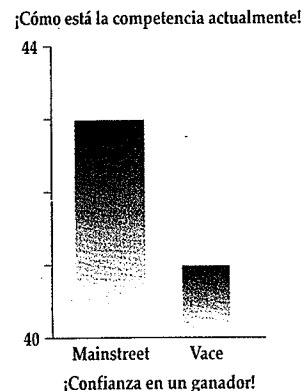


gráfico en manos del individuo anónimo que diseñó el programa. En consecuencia, los medios de comunicación masiva (al contrario de los estrictamente científicos) nos abruma con gráficos generados de manera instantánea por la computadora. Cuando usted llegue a ser experto en el pensamiento estadístico, empezará a notar que muchos, si no la mayoría de estos "instantáneos", son poco confiables en el mejor de los casos (es decir, abiertos a múltiples interpretaciones), y en el peor, incorrectos.

La siguiente parábola ilustra una distorsión gráfica común. El candidato (ficticio) a gobernador Mortimer Mainstreet tenía un cómodo margen de dos a uno en las primeras encuestas electorales sobre su único serio contendiente, Harry Vace. Poco después, sin embargo, su delantera empezó a revertirse. Se rumora que Mainstreet está a punto de despedir a su personal de campaña, cuyos miembros ya temen que sus sueños de gobernar el estado se desvanezcan.

La última encuesta señala que la ventaja de Mortimer disminuyó 2 puntos porcentuales, de 43 a 41 por ciento, con el 16 por ciento aún indeciso y un margen de error de más menos 3 por ciento. La competencia se ha vuelto muy cerrada. En un intento por evitar perder sus empleos, los miembros del personal de campaña informan a Mortimer las puntuaciones que muestra la figura 3-8 (¿Si Mortimer Mainstreet se deja llevar por esto, no merece ser gobernador! ¿Puedes identificar todas las cuestiones erróneas en este gráfico?

RESUMEN

1. Las tablas y gráficos dan un sentido de proporción acerca de una distribución de puntuaciones sin que haya necesidad de que el lector tenga profundos conocimientos de matemáticas.
2. Selecciona un diseño gráfico con base en a) el nivel de medición de la variable, b) los objetivos del estudio, y c) la audiencia a quien va dirigido.
3. Hay lineamientos razonados para construir gráficos y tablas. Un gráfico debe simplificar, no complicar. Una tabla debe explicarse por sí misma y tener sentido, sin tener que referirse a un texto.
4. Generalmente se emplean gráficos de pastel y gráficos de barras para ilustrar la distribución de categorías de una variable nominal/ordinal.
5. Los gráficos de pastel son especialmente útiles para expresar un sentido de claridad, tamaño relativo o desigualdad entre categorías.
6. Los gráficos de barras son especialmente útiles para expresar un sentido de competencia entre categorías.
7. Los gráficos de barras agrupadas son buenos para comparar dos o más grupos para una variable nominal/ordinal.
8. Para variables de intervalo/razón, utiliza un histograma de frecuencia o polígono de frecuencia (gráfico de línea).
9. Los histogramas de frecuencias atraen la atención hacia donde cae el grueso de puntuaciones en una distribución.
10. Los polígonos de frecuencias (o gráficos de línea) describen un sentido de tendencia o movimiento en una distribución de puntuaciones.