

AGRUPACIÓN Y PRESENTACIÓN DE DATOS PARA EXPRESAR SIGNIFICADOS: TABLAS Y GRÁFICAS

Objetivos

- Mostrar la diferencia entre muestras y poblaciones
- Convertir los datos sin procesar en información útil
- Construir y utilizar datos ordenados
- Construir y utilizar distribuciones de frecuencias
- Representar gráficamente distribuciones de frecuencias con histogramas, polígonos y ojivas
- Utilizar distribuciones de frecuencias para tomar decisiones

Contenido del capítulo

- 2.1 ¿Cómo podemos ordenar los datos? 8
- 2.2 Ejemplos de datos sin procesar 11
- 2.3 Ordenamiento de datos en arreglos de datos y distribuciones de frecuencias 12
- 2.4 Construcción de una distribución de frecuencias 20
- 2.5 Representación gráfica de distribuciones de frecuencias 29
- Estadística en el trabajo 42
- Ejercicio de base de datos computacional 43
- Términos introducidos en el capítulo 2 45
- Ecuaciones introducidas en el capítulo 2 46
- Ejercicios de repaso 46



El gerente de producción de la Dalmon Carpet Company es responsable de la fabricación de alfombras en más de 500 telares. Para no tener que medir la producción diaria (en yardas) de cada telar, toma una muestra de 30 telares cada día, con lo que llega a una conclusión respecto a la producción promedio de alfombras de las 500 máquinas. La tabla que se presenta más adelante exhibe la producción de cada uno de los 30 telares de la muestra. Estas cantidades son los datos sin procesar a partir de los cuales el director de producción puede sacar conclusiones acerca de la toda la población de telares en su desempeño del día anterior.

Producción en yardas de 30 telares para alfombra

16.2	15.4	16.0	16.6	15.9	15.8	16.0	16.8	16.9	16.8
15.7	16.4	15.2	15.8	15.9	16.1	15.6	15.9	15.6	16.0
16.4	15.8	15.7	16.2	15.6	15.9	16.3	16.3	16.0	16.3

Mediante los métodos introducidos en este capítulo, podemos ayudar al director de producción a llegar a la conclusión correcta. ■

Algunas definiciones Los *datos* son colecciones de cualquier cantidad de observaciones relacionadas. Podemos recopilar el número de teléfonos que diferentes empleados instalan en un día dado o el número de teléfonos que instala un trabajador dado durante un día en un periodo de varios días, y podemos llamar datos a estos resultados. Una colección de datos se conoce como *conjunto de datos*; una sola observación es un *dato puntual*.

2.1 ¿Cómo podemos ordenar los datos?

Para que los datos sean útiles, necesitamos organizar nuestras observaciones de modo que podamos distinguir patrones y llegar a conclusiones lógicas. Este capítulo presenta las técnicas tabular y gráfica para organización de datos; en el capítulo 3 mostraremos cómo usar números para describir datos.

Recolección de datos

Represente todos los grupos

Los especialistas en estadística seleccionan sus observaciones de manera que todos los grupos relevantes estén representados en los datos. Para determinar el mercado potencial de un nuevo producto, por ejemplo, los analistas podrían estudiar cien consumidores de cierta área geográfica. Dichos analistas deben tener la certeza de que este grupo incluya personas que representan distintas variables como nivel de ingresos, raza, nivel educativo y vecindario.

Encuentre datos a partir de observaciones o de registros

Los datos pueden provenir de observaciones reales o de registros elaborados con otros propósitos. Por ejemplo, con fines de facturación e informes médicos, un hospital registra el número de pacientes que utilizan los servicios de rayos X. Esta información puede organizarse para producir datos que los especialistas en estadística puedan describir e interpretar.

Use los datos del pasado para tomar decisiones del futuro

Los datos pueden ayudar a los responsables de tomar decisiones a hacer suposiciones razonadas acerca de las *causas* y, por tanto, de los *efectos* probables de ciertas características en situaciones dadas. El conocimiento de tendencias derivado de la experiencia previa puede, también, permitir a los ciudadanos conscientes anticipar resultados posibles y actuar en consecuencia. Una investigación de mercado puede revelar que determinado producto es preferido por amas de casa de origen afroamericano de las comunidades suburbanas, con ingresos y educación promedios. Los textos publicitarios del producto deberán, por tanto, estar dirigidos a esta audiencia. Y si los registros del hospital mues-

tran que un número mayor de pacientes ha utilizado el aparato de rayos X en junio que en enero, el departamento de personal del hospital deberá determinar si esto sucedió sólo ese año o es indicativo de una tendencia y, tal vez, debería ajustar su programa de contratación y asignación de vacaciones de acuerdo con la información.

Cuando los datos se ordenan de manera compacta y útil, los responsables de tomar decisiones pueden obtener información confiable sobre el entorno y usarla para tomar decisiones inteligentes. En la actualidad, las computadoras permiten a los especialistas en estadística recolectar enormes volúmenes de observaciones y comprimirlas en tablas, gráficas y cifras instantáneamente. Éstas son formas compactas y útiles, pero ¿son confiables? Recuerde que los datos producidos por una computadora son tan precisos como los datos que entraron en ella. Como dicen los programadores, ¡“BEBS”! o ¡“basura entra, basura sale!”!. Los administradores deben tener mucho cuidado y cerciorarse que los datos empleados estén basados en suposiciones e interpretaciones correctas. Antes de depositar nuestra confianza en cualquier conjunto de datos interpretados, vengan de una computadora o no, póngalos a prueba mediante las siguientes preguntas:

Pruebas para datos

1. ¿De dónde vienen los datos? ¿La fuente es tendenciosa?, es decir, ¿es posible que exista interés en proporcionar datos que conduzcan a cierta conclusión más que a otras?
2. ¿Los datos apoyan o contradicen otras evidencias que se tienen?
3. ¿Hace falta alguna evidencia cuya ausencia podría ocasionar que se llegue a una conclusión diferente?
4. ¿Cuántas observaciones se tienen? ¿Representan a todos los grupos que se desea estudiar?
5. ¿La conclusión es lógica? ¿Se ha llegado a conclusiones que los datos no confirman?

Estudie las respuestas que dé a estas preguntas. ¿Vale la pena usar los datos o debemos esperar y recabar más información antes de actuar? Si el hospital se queda corto de personal porque contrató pocos técnicos para operar las instalaciones de rayos X, quiere decir que su administrador confió en datos insuficientes. Si la agencia de publicidad dirigió su propaganda sólo a las amas de casa afroamericanas de los suburbios cuando pudo haber triplicado sus ventas al incluir también a amas de casa blancas, quiere decir que también en este caso hubo confianza en datos insuficientes. En ambos casos, probar los datos disponibles podría haber ayudado a los administradores a tomar mejores decisiones.

Ejemplo de conteo doble

El efecto de emplear datos incompletos o tendenciosos puede ilustrarse con el siguiente ejemplo: una asociación nacional de líneas de transporte afirma, en un comercial, que “el 75% de todo lo que uno utiliza se transporta en camión”. Esta afirmación podría conducir a creer que automóviles, ferrocarriles, aeroplanos, barcos y otras formas de transporte sólo trasladan el 25% de lo que utilizamos. Es fácil llegar a tal conclusión, pero ésta no es completamente ilustrativa. En la afirmación no se toma en cuenta la cuestión del “conteo doble”. ¿Qué pasa cuando algo es llevado hasta su ciudad por ferrocarril y luego en camión hasta su casa? ¿O cómo se considera la paquetería si se transporta por correo aéreo y luego en motocicleta? Cuando se resuelve la cuestión del conteo doble (algo bastante complicado de manejar), resulta que los camiones transportan una proporción mucho menor de bienes que la afirmada por los transportistas. Aunque los camiones están involucrados en la *entrega* de una proporción relativamente alta de lo que utilizamos, el ferrocarril y los barcos llevan más bienes a lo largo de más kilómetros.

Diferencia entre muestras y poblaciones

Definición de muestra y población

Los expertos en estadística recogen datos de una muestra y utilizan esta información para hacer inferencias sobre la población que representa esa muestra. Así, una población es un todo y una muestra es una fracción o segmento de ese todo.

Función de las muestras

Se estudiarán las muestras para poder describir las poblaciones. El hospital del ejemplo puede estudiar un grupo pequeño y representativo de los registros de rayos X, en lugar de examinar cada uno de los registros de los últimos 50 años. La Gallup Poll puede entrevistar una muestra de sólo 2,500

estadounidenses adultos con el fin de predecir la opinión de todos los adultos que viven en Estados Unidos.

Ventajas de las muestras

El estudio de una muestra es más sencillo que el de la población completa, cuesta menos y lleva menos tiempo. A menudo, probar la resistencia de una parte de avión implica destruirla; en consecuencia, es deseable probar la menor cantidad de partes. En algunas ocasiones, la prueba implica un riesgo humano; el uso de muestras disminuye ese riesgo a un nivel aceptable. Por último, se ha probado que incluso el examen de una población entera deja pasar algunos elementos defectuosos. Por tanto, en algunos casos, el muestreo puede *eleva*r el nivel de calidad. Si usted se pregunta cómo puede suceder esto, piense en lo cansado y poco animoso que estaría si tuviera que observar de manera continua miles y miles de productos en una banda continua.

Función de las poblaciones

Una *población* es un conjunto de todos los elementos que estamos estudiando, acerca de los cuales intentamos sacar conclusiones. Debemos definir esa población de modo que quede claro cuándo cierto elemento pertenece o no a la población. Para el estudio de mercado mencionado, la población puede estar constituida por todas las mujeres residentes dentro de un radio de 25 kilómetros con centro en la ciudad de Cincinnati, que tengan un ingreso familiar anual entre \$20,000 y \$45,000 y hayan cursado por lo menos 11 años de educación formal. Una mujer que viva en el centro de Cincinnati, con ingresos familiares de \$25,000 y estudios universitarios completos podría ser parte de esa población, mientras que una mujer residente de San Francisco, o que tenga ingresos familiares de \$7,000 o con cinco años de estudios profesionales, no se consideraría miembro de nuestra población.

Necesidad de una muestra representativa

Una *muestra* es una colección de algunos elementos de la población, no de todos. La población de nuestra investigación de mercado son *todas* las mujeres que cumplen con los requisitos anteriores. Cualquier grupo de mujeres que cumpla con tales requisitos puede constituir una muestra, siempre y cuando el grupo sea una fracción de la población completa. Una porción grande de relleno de frambuesas con sólo algunas migajas de corteza es una muestra de tarta, pero no es una muestra representativa debido a que las proporciones de los ingredientes no son las mismas en la muestra que en el todo.

Una *muestra representativa* contiene las características relevantes de la población *en las mismas proporciones* en que están incluidas en tal población. Si la población de mujeres del ejemplo es un tercio de afroamericanas, entonces una muestra representativa en términos de raza tendrá también un tercio de afroamericanas. En el capítulo 6 se analizan con más detalle los métodos específicos para realizar muestreos.

Búsqueda de un patrón significativo en los datos

Los datos pueden organizarse de muchas maneras

Los datos pueden organizarse de muchas maneras. Podemos, simplemente recolectarlos y mantenerlos en orden. O, si las observaciones están hechas con números, podemos hacer una lista de los datos puntuales de menor a mayor según su valor numérico. Pero si los datos consisten en trabajadores especializados (como carpinteros, albañiles o soldadores) de una construcción, o en los distintos tipos de automóviles que ensamblan todos los fabricantes, o en los diferentes colores de suéteres fabricados por una empresa dada, debemos organizarlos de manera distinta. Necesitaremos presentar los datos puntuales en orden alfabético o mediante algún principio de organización. Una forma común de organizar los datos consiste en dividirlos en categorías o clases parecidas y luego contar el número de observaciones que quedan dentro de cada categoría. Este método produce una *distribución de frecuencias*, que se estudiará más adelante, en este mismo capítulo.

¿Por qué debemos organizar los datos?

Organizar los datos tiene como fin permitirnos ver rápidamente algunas de las características de los datos recogidos. Buscamos cosas como el alcance (los valores mayor y menor), patrones evidentes, alrededor de qué valores tienden a agruparse los datos, qué valores aparecen con más frecuencia, etc. Cuanta más información de este tipo podamos obtener de una muestra, mejor entenderemos la población de la cual proviene, y mejor será nuestra toma de decisiones.

Ejercicios 2.1

Aplicaciones

- **2-1** Cuando se les pregunta qué medicamento tomarían si estuvieran en una isla abandonada y sólo pudieran elegir un analgésico, la mayoría de los médicos prefiere Bayer, en lugar de Tylenol, Bufferin o Advil. ¿Se llegó a esta conclusión a partir de una muestra o de una población?
- **2-2** El 25% de los automóviles vendidos en Estados Unidos en 1992 fue armado en Japón. ¿Se llegó a esta conclusión a partir de una muestra o de una población?
- **2-3** Hace poco, una fábrica de aparatos electrónicos introdujo al mercado un nuevo amplificador, y en la garantía se indica que se han vendido 10,000 de esos amplificadores. El presidente de la empresa, bastante preocupado después de recibir tres cartas de quejas con respecto a los nuevos amplificadores, informó al gerente de producción que se implantarían costosas medidas de control para asegurar que los defectos no aparecieran de nuevo. Haga un comentario sobre la reacción del gerente desde la perspectiva de las cinco pruebas para datos que se dieron en la página 9.
- **2-4** “Alemania permanecerá siempre dividida”, afirmó Walter Ulbricht después de la construcción del Muro de Berlín en 1961. Sin embargo, hacia finales de 1969, el gobierno de Alemania del Este empezó a permitir el libre tránsito entre la parte oriental y la occidental, y veinte años después, el muro fue destruido por completo. Mencione algunas razones por las cuales Ulbricht hizo una predicción incorrecta.
- **2-5** Analice los datos dados en el problema planteado al inicio del capítulo en términos de las cinco pruebas para datos dados en la página 9.

2.2 Ejemplos de datos sin procesar

La información obtenida, antes de ser organizada y analizada, se conoce como *datos sin procesar*. Son “sin procesar” puesto que aún no han sido manejados mediante un método estadístico.

Problema que enfrenta el personal de admisión

Los datos de los telares proporcionados al inicio del capítulo son un ejemplo de datos sin procesar. Consideremos un segundo ejemplo. Suponga que el personal de admisión de una universidad, preocupado por el éxito de los estudiantes que escogen para ser admitidos, desea comparar el desempeño de los estudiantes universitarios con otros logros, como calificaciones en el nivel anterior, resultados de exámenes y actividades extracurriculares. En lugar de estudiar a todos los alumnos de todos los años, el personal puede tomar una muestra de la población de estudiantes en un periodo dado y estudiar sólo ese grupo, para llegar a conclusiones sobre qué características parecen predecir el éxito. Por ejemplo, puede comparar las calificaciones del nivel educativo anterior con el promedio de calificaciones universitarias de los estudiantes de la muestra. El personal puede sumar las calificaciones y dividir el resultado entre el número total de ellas para obtener un promedio, o media, para cada estudiante. En la tabla 2-1 se presentan tabulados estos datos sin procesar: 20 pares de calificaciones promedio en el bachillerato y en la universidad.

Problema de construcción de un puente

Cuando se diseña un puente, los ingenieros deben determinar la tensión que un material dado, digamos el concreto, puede soportar. En lugar de probar cada pulgada cúbica de concreto para determinar su capacidad de resistencia, los ingenieros toman una muestra, la prueban y llegan a la conclusión sobre qué tanta tensión, en promedio, puede resistir ese tipo de concreto. En la tabla 2-2 se presentan los datos sin procesar obtenidos de una muestra de 40 bloques de concreto que se utilizarán para construir un puente.

Tabla 2-1	Bach.	Univ.	Bach.	Univ.	Bach.	Univ.	Bach.	Univ.
Promedios de calificaciones en bachillerato y universidad de 20 estudiantes universitarios del último año	3.6	2.5	3.5	3.6	3.4	3.6	2.2	2.8
	2.6	2.7	3.5	3.8	2.9	3.0	3.4	3.4
	2.7	2.2	2.2	3.5	3.9	4.0	3.6	3.0
	3.7	3.2	3.9	3.7	3.2	3.5	2.6	1.9
	4.0	3.8	4.0	3.9	2.1	2.5	2.4	3.2

Presión en libras por pulgada cuadrada que puede resistir el concreto	2500.2	2497.8	2496.9	2500.8	2491.6	2503.7	2501.3	2500.0
	2500.8	2502.5	2503.2	2496.9	2495.3	2497.1	2499.7	2505.0
	2490.5	2504.1	2508.2	2500.8	2502.2	2508.1	2493.8	2497.8
	2499.2	2498.3	2496.7	2490.4	2493.4	2500.7	2502.0	2502.5
	2506.4	2499.9	2508.4	2502.3	2491.3	2509.5	2498.4	2498.1

SUGERENCIAS Y SUPOSICIONES Los datos no necesariamente son información; tener más datos no necesariamente produce mejores decisiones. La meta es resumir y presentar los datos de manera útil para apoyar la toma de decisiones efectiva y ágil. La razón por la que los datos deben organizarse es ver si existe un patrón en ellos, patrones como el valor más grande y el más pequeño, o el valor alrededor del cual parecen agruparse. Si los datos provienen de una muestra, se suponen representativos de la población de la que se tomaron. Todos los buenos estadísticos (y usuarios de datos) reconocen que usar datos sesgados o incompletos conduce a malas decisiones.

Ejercicios 2.2

Aplicaciones

- **2-6** Observe los datos de la tabla 2-1. ¿Por qué los datos necesitan organizarse más? ¿Puede usted llegar a una conclusión partiendo de los datos tal como se presentan?
- **2-7** El gerente de mercadotecnia de una compañía grande recibe un informe mensual de las ventas de uno de los productos de la compañía. El informe consiste en una lista de las ventas del producto por estado durante el mes anterior. ¿Es éste un ejemplo de datos sin procesar?
- **2-8** El gerente de producción de una compañía grande recibe un informe mensual de su departamento de control de calidad. El informe presenta el índice de rechazo de la línea de producción (número de rechazos por cada 100 unidades producidas), la máquina que ocasiona el mayor número de productos defectuosos y el costo promedio de reparación de las unidades rechazadas. ¿Es éste un ejemplo de datos sin procesar?

2.3 Ordenamiento de datos en arreglos de datos y distribuciones de frecuencias

Definición de ordenación de datos

Una *ordenación de datos* es una de las formas más sencillas de presentarlos: organiza los valores en orden ascendente o descendente. En la tabla 2-3 repetimos los datos de la producción de alfombras del problema presentado al inicio del capítulo, y en la tabla 2-4 reordenamos los números en un orden ascendente.

Ventajas de la ordenación de datos

La ordenación de datos ofrece varias ventajas con respecto a los datos sin procesar:

1. **Podemos identificar los valores mayor y menor rápidamente.** En el ejemplo de las alfombras, los datos van de las 15.2 a las 16.9 yardas.

Muestra de la producción diaria de 30 telares para alfombra (en yardas)	16.2	15.8	15.8	15.8	16.3	15.6
	15.7	16.0	16.2	16.1	16.8	16.0
	16.4	15.2	15.9	15.9	15.9	16.8
	15.4	15.7	15.9	16.0	16.3	16.0
	16.4	16.6	15.6	15.6	16.9	16.3

Tabla 2-4						
Ordenamiento de los datos de la producción diaria de 30 telares para alfombra (en yardas)	15.2	15.7	15.9	16.0	16.2	16.4
	15.4	15.7	15.9	16.0	16.3	16.6
	15.6	15.8	15.9	16.0	16.3	16.8
	15.6	15.8	15.9	16.1	16.3	16.8
	15.6	15.8	16.0	16.2	16.4	16.9

- 2. Es fácil dividir los datos en secciones.** En la tabla 2-4, los primeros 15 valores (la mitad inferior de los datos) están entre 15.2 y 16.0 yardas, y los últimos 15 (la mitad superior) entre 16.0 y 16.9 yardas. De manera similar, el tercio más bajo de los datos está entre 15.2 y 15.8 yardas, el tercio de en medio está entre 15.9 y 16.2 yardas, y el tercio más alto entre 16.2 y 16.9 yardas.
- 3. Podemos ver si algunos valores aparecen más de una vez en el arreglo.** Los valores iguales quedan juntos. En la tabla 2-4 se muestra que nueve niveles ocurren más de una vez cuando se toma una muestra de 30 telares.
- 4. Podemos observar la distancia entre valores sucesivos de los datos.** En la tabla 2-4, 16.6 y 16.8 son valores sucesivos. La distancia entre ellos es de 0.2 yardas ($16.8 - 16.6$).

Desventajas de la ordenación de datos

A pesar de las ventajas, en algunas ocasiones un ordenamiento de datos no resulta útil. Como da una lista de todos los valores, es una forma incómoda de mostrar grandes cantidades de datos. Suele ser necesario comprimir la información sin que pierda su utilidad para la interpretación y la toma de decisiones. ¿Cómo podemos hacerlo?

Una mejor forma de organizar datos: la distribución de frecuencias

Las distribuciones de frecuencias permiten manejar más datos

Una manera de compactar datos es mediante una *tabla de frecuencias* o *distribución de frecuencias*. Para entender la diferencia entre ésta y el ordenamiento de datos, tome como ejemplo las existencias promedio (en días) para 20 tiendas de autoservicio:

Tabla 2-5					
Ordenamiento de datos para las existencias promedio (en días) para 20 tiendas de autoservicio	2.0	3.8	4.1	4.7	5.5
	3.4	4.0	4.2	4.8	5.5
	3.4	4.1	4.3	4.9	5.5
	3.8	4.1	4.7	4.9	5.5

Pierden algo de información

Pero se obtiene otro tipo de información

En las tablas 2-5 y 2-6, hemos tomado los mismos datos concernientes a las existencias promedio y los hemos organizado, primero, como un arreglo ascendente y luego como una distribución de frecuencias. Para obtener la tabla 2-6 tuvimos que dividir los datos en grupos de valores parecidos. Después registramos el número de datos puntuales que caen en cada grupo. Observe que perdimos algo de información al construir la distribución de frecuencias. Ya no podemos saber, por ejemplo, que el valor 5.5 aparece cuatro veces y que el valor 5.1 no aparece. Sin embargo, ganamos información acerca del *patrón* de existencias promedio. En la tabla 2-6, podemos ver que las existencias promedio caen con más frecuencia en el intervalo de 3.8 a 4.3 días. No es usual encontrar existencias

Tabla 2-6	Clase (grupo de valores parecidos de datos puntuales)	Frecuencia (número de observaciones que hay en cada clase)
Distribución de frecuencias para las existencias promedio (en días) en 20 tiendas de autoservicio (6 clases)	2.0 a 2.5	1
	2.6 a 3.1	0
	3.2 a 3.7	2
	3.8 a 4.3	8
	4.4 a 4.9	5
	5.0 a 5.5	4

promedio en el intervalo de 2.0 a 2.5 días o en el que va de 2.6 a 3.1 días. Las existencias que están entre 4.4 y 4.9 días y entre 5.0 y 5.5 días no son las que prevalecen, pero ocurren con más frecuencia que algunas otras. Así pues, las distribuciones de frecuencias sacrifican algunos detalles pero nos ofrecen nuevas perspectivas sobre los patrones de los datos.

Función de las clases en una distribución de frecuencias

Una distribución de frecuencias es una tabla en la que organizamos los datos en clases, es decir, en grupos de valores que describen una característica de los datos. El inventario promedio es una característica de las 20 tiendas de autoservicio. En la tabla 2-5, esta característica tiene once valores diferentes. Pero estos mismos datos podrían dividirse en cualquier número de clases. En la tabla 2-6, por ejemplo, utilizamos seis. Podríamos comprimir todavía más los datos y utilizar sólo dos clases: menores que 3.8 y mayores o iguales que 3.8. O podríamos aumentar el número de clases utilizando intervalos más pequeños, como se hace en la tabla 2-7.

¿Por qué se le conoce como distribución de frecuencias?

Una distribución de frecuencias muestra el **número de observaciones del conjunto de datos que caen en cada una de las clases**. Si se puede determinar la frecuencia con la que se presentan los valores en cada clase de un conjunto de datos, se puede construir una distribución de frecuencias.

Definición de distribución de frecuencias relativas

Características de las distribuciones de frecuencias relativas

Hasta aquí hemos considerado la frecuencia con que aparecen los valores en cada clase como el número total de datos puntuales u observaciones que caen en cada clase. Podemos expresar la frecuencia de cada valor también como una *fracción* o un *porcentaje* del número total de observaciones. La frecuencia de un inventario promedio de 4.4 a 4.9 días, por ejemplo, es 5 en la tabla 2-6, pero 0.25 en la tabla 2-8. Para obtener este valor de 0.25, dividimos la frecuencia de esa clase (5) entre el número total de observaciones del conjunto de datos (20). La respuesta se puede expresar como una fracción ($\frac{5}{20}$), un número decimal (0.25) o un porcentaje (25%). En una *distribución de frecuencias relativas* se presentan las frecuencias en términos de fracciones o porcentajes.

Las clases son completamente incluyentes

Observe, en la tabla 2-8, que la suma de todas las frecuencias relativas es igual a 1.00, o 100%. Esto se debe a que una distribución de frecuencias relativas da una correspondencia de cada clase con su fracción o porcentaje del total de los datos. Por consiguiente, las clases que aparecen en cualquier distribución de frecuencias, sean relativas o simples, son *completamente incluyentes*. Todos los datos caen en una u otra categoría. Note también que las clases de la tabla 2-8 son *mutuamente excluyentes*, es decir, ningún dato puntual cae en más de una categoría. En la tabla 2-9 se ilustra este concepto mediante la comparación de clases mutuamente excluyentes con clases que se traslapan. En las distribuciones de frecuencias no existen clases que se traslapan.

Son mutuamente excluyentes

Tabla 2-7	Clase	Frecuencia	Clase	Frecuencia
Distribución de frecuencias para las existencias promedio (en días) en 20 tiendas de autoservicio (12 clases)	2.0 a 2.2	1	3.8 a 4.0	3
	2.3 a 2.5	0	4.1 a 4.3	5
	2.6 a 2.8	0	4.4 a 4.6	0
	2.9 a 3.1	0	4.7 a 4.9	5
	3.2 a 3.4	2	5.0 a 5.2	0
	3.5 a 3.7	0	5.3 a 5.5	4

Tabla 2-8	Clase	Frecuencia	Frecuencia relativa: fracción de observaciones en cada clase
Distribución de frecuencias relativas del inventario promedio (en días) para 20 tiendas de autoservicio	2.0 a 2.5	1	0.05
	2.6 a 3.1	0	0.00
	3.2 a 3.7	2	0.10
	3.8 a 4.3	8	0.40
	4.4 a 4.9	5	0.25
	5.0 a 5.5	4	0.20
		<u>20</u>	<u>1.00</u>

Tabla 2-9					
Clases mutuamente excluyentes y clases que se traslapan	Mutuamente excluyentes	1 a 4	5 a 8	9 a 12	13 a 16
	No mutuamente excluyentes	1 a 4	3 a 6	5 a 8	7 a 10

Clases de datos cualitativos

Hasta ahora, las clases han consistido en números y describen algún atributo cuantitativo de los elementos de la muestra. Podemos, también, clasificar la información de acuerdo con características cualitativas, como raza, religión y sexo, que no entran de manera natural en categorías numéricas. Igual que las clases de atributos cuantitativos, éstas deben ser completamente incluyentes y mutuamente excluyentes. En la tabla 2-10 se muestra cómo construir distribuciones de frecuencias simples o absolutas y distribuciones de frecuencias relativas usando el atributo cualitativo de ocupación.

Clases de extremo abierto para listas que no son exhaustivas

Aunque en la tabla 2-10 no se enumeran todas las ocupaciones de los graduados del *Central College*, sigue siendo completamente incluyente. ¿Por qué? La clase “otros” cubre a todas las observaciones que no entran en las categorías mencionadas. Utilizaremos algo parecido siempre que nuestra lista no incluya específicamente todas las posibilidades. Por ejemplo, si la característica puede presentarse durante cualquier mes del año, una lista completa debería incluir 12 categorías. Pero si deseamos enumerar sólo los ocho primeros meses, de enero a agosto, podemos utilizar el término *otros* para referirnos a las observaciones correspondientes a los restantes cuatro meses, septiembre, octubre, noviembre y diciembre. Aunque nuestra lista no incluye específicamente todas las posibilidades, sigue siendo completamente incluyente. La categoría “otros” se conoce como *clase de extremo abierto* cuando permite que el extremo inferior o el superior de una clasificación cuantitativa no esté limitado. La última clase de la tabla 2-11 (“72 o más”) es de extremo abierto.

Tabla 2-10			
Ocupación de los integrantes de una muestra de 100 graduados del <i>Central College</i>	Clase ocupacional	Distribución de frecuencias (1)	Distribución de frecuencias relativas (1) ÷ 100
	Actor	5	0.05
	Banquero	8	0.08
	Empresario	22	0.22
	Químico	7	0.07
	Médico	10	0.10
	Agente de seguros	6	0.06
	Periodista	2	0.02
	Abogado	14	0.14
	Maestro	9	0.09
	Otros	17	0.17
		<u>100</u>	<u>1.00</u>

Tabla 2-11			
Edades de los habitantes del condado de Bunder	Clase: edad (1)	Frecuencia (2)	Frecuencia relativa (2) ÷ 89,592
	Nacimiento a 7	8,873	0.0990
	8 a 15	9,246	0.1032
	16 a 23	12,060	0.1346
	24 a 31	11,949	0.1334
	32 a 39	9,853	0.1100
	40 a 47	8,439	0.0942
	48 a 55	8,267	0.0923
	56 a 63	7,430	0.0829
	64 a 71	7,283	0.0813
	72 o más	6,192	0.0691
		<u>89,592</u>	<u>1.0000</u>

Clases discretas

Los esquemas de clasificación pueden ser tanto cuantitativos como cualitativos y tanto discretos como continuos. Las *clases discretas* son entidades separadas que no pasan de una clase a la siguiente sin que haya un rompimiento. Clases como el número de niños de cada familia, el número de camiones de una compañía de transportistas o las ocupaciones de los graduados del *Central College* son discretas. Los datos discretos son aquellos que pueden tomar sólo un número limitado de valores o un número infinito numerable de valores. Los graduados del *Central College* pueden clasificarse como doctores o químicos, pero no como algo intermedio. El precio de cierre de las acciones de AT&T puede ser $39\frac{1}{2}$ o $39\frac{7}{8}$ (pero no 39.43), o su equipo de básquetbol favorito puede ganar por 5 o 27 puntos (pero no por 17.6 puntos).

Clases continuas

Los *datos continuos* pasan de una clase a otra sin que haya un rompimiento. Implican mediciones numéricas como el peso de las latas de tomates, la presión sobre el concreto o las calificaciones de bachillerato de los estudiantes del último año en la universidad, por ejemplo. Los datos continuos pueden expresarse con números fraccionarios o con enteros.

SUGERENCIAS Y SUPOSICIONES Existen muchas maneras de presentar datos. Para empezar, puede elaborar un ordenamiento de datos de manera descendente o ascendente. Mostrar cuántas veces aparece un valor usando una distribución de frecuen-

cias es aún más efectivo; convertir estas frecuencias en decimales (que se conocen como frecuencias relativas) puede ayudar todavía más. Sugerencia: debe recordar que las variables discretas son cosas que se pueden contar y las variables continuas aparecen en algún punto de una escala.

Ejercicios 2.3

Ejercicios de autoevaluación

EA 2-1 Las edades de los 50 integrantes de un programa de servicio social del gobierno son:

83	51	66	61	82	65	54	56	92	60
65	87	68	64	51	70	75	66	74	68
44	55	78	69	98	67	82	77	79	62
38	88	76	99	84	47	60	42	66	74
91	71	83	80	68	65	51	56	73	55

Use estos datos para construir las distribuciones de frecuencia relativa con 7 y 13 intervalos iguales. La política del estado para los programas de servicio social requiere que alrededor de 50% de los participantes tengan más de 50 años.

- ¿Cumple el programa con la política?
- ¿La distribución de frecuencias relativas de 13 intervalos ayuda a responder el inciso a) mejor que la distribución de 7 intervalos?
- Suponga que el director de servicios sociales desea saber la proporción de participantes en el programa que tienen entre 45 y 50 años de edad. ¿A partir de cuál distribución de frecuencias relativas, de 7 o de 13 intervalos, puede estimar mejor la respuesta?

EA 2-2 Use los datos de la tabla 2-1 para elaborar un ordenamiento de mayor a menor del promedio general en bachillerato. Después haga un ordenamiento de mayor a menor del promedio general en la universidad. ¿A partir de los dos ordenamientos, qué puede concluir que no podía a partir de los datos originales?

Aplicaciones

- 2-9** Los talleres Transmissions Fix-It registran el número de comprobantes de servicio extendidos el mes anterior en cada una de sus 20 sucursales de la forma siguiente:

823	648	321	634	752
669	427	555	904	586
722	360	468	847	641
217	588	349	308	766

La compañía tiene la creencia de que una sucursal no puede mantenerse financieramente con menos de 475 servicios mensuales. Es también política de la compañía otorgar una bonificación económica al gerente de la sucursal que genere más de 725 servicios mensuales. Ordene los datos de la tabla e indique cuántas sucursales no pueden mantenerse y cuántas recibirán bonificación.

- **2-10** Utilice los datos de la empresa Transmissions Fix-It del ejercicio 2-9. La vicepresidente financiera de la compañía ha establecido lo que llama una “lista de observación de sucursales” que contiene las sucursales cuya actividad en cuanto a servicios prestados es lo suficientemente baja como para que la casa matriz le preste atención especial. Esta categoría incluye a las sucursales cuya actividad está entre 550 y 650 servicios mensuales. ¿Cuántos talleres deberán estar en dicha lista si nos basamos en la actividad del mes anterior?
- **2-11** El número de horas que les toma a los mecánicos retirar, reparar y reinstalar una transmisión en uno de los talleres de Transmissions Fix-It, durante un día de la semana anterior, se registra de la manera siguiente:

4.3	2.7	3.8	2.2	3.4
3.1	4.5	2.6	5.5	3.2
6.6	2.0	4.4	2.1	3.3
6.3	6.7	5.9	4.1	3.7

A partir de estos datos, elabore una distribución de frecuencias con intervalos de una hora. ¿A qué conclusiones puede llegar acerca de la productividad de los mecánicos si toma en cuenta la distribución de frecuencias? Si el gerente de la Transmissions Fix-It cree que más de 6.0 horas es evidencia de un desempeño insatisfactorio, ¿de qué magnitud es el problema del desempeño de los mecánicos en este taller en particular?

- **2-12** El comisionado de transporte del condado de Orange está preocupado por la velocidad a la que los conductores manejan en un tramo de la carretera principal. Los datos de velocidad de 45 conductores son los siguientes:

15	32	45	46	42	39	68	47	18
31	48	49	56	52	39	48	69	61
44	42	38	52	55	58	62	58	48
56	58	48	47	52	37	64	29	55
38	29	62	49	69	18	61	55	49

Use estos datos para elaborar distribuciones de frecuencias relativas con 5 y 11 intervalos iguales. El Departamento de Transporte informa que, a nivel nacional, no más de 10% de los conductores excede 55 mph.

- a) ¿Se comportan los conductores del condado de Orange de acuerdo con las afirmaciones del informe del Departamento de Transporte acerca de los patrones de manejo?
- b) ¿Qué distribución usó para responder el inciso a)?
- c) El Departamento de Transporte ha determinado que la velocidad más segura para esta carretera es más de 36 y menos de 59 mph. ¿Qué proporción de conductores maneja dentro de este intervalo? ¿Qué distribución ayuda a responder a esta pregunta?
- **2-13** Ordene los datos de la tabla 2-2, en un arreglo de mayor a menor.
 - a) Suponga que la ley estatal requiere que los puentes de concreto puedan soportar al menos 2,500 lb/pulg². ¿Cuántas muestras no pasarán esta prueba?
 - b) ¿Cuántas muestras podrían soportar una presión de al menos 2,497 lb/pulg² pero no una de 2,504 lb/pulg²?
 - c) Si examina con cuidado el arreglo, se dará cuenta de que algunas muestras pueden soportar cantidades iguales de presión. Proporcione una lista de tales presiones y el número de muestras que pueden soportar cada cantidad de presión.
- **2-14** Un estudio reciente sobre los hábitos de los consumidores de televisión por cable en Estados Unidos proporcionó los siguientes datos:

Número de canales comprados	Número de horas que ven televisión por semana
25	14
18	16

Continúa

Número de canales comprados	Número de horas que ven televisión por semana
42	12
96	6
28	13
43	16
39	9
29	7
17	19
84	4
76	8
22	13
104	6

Ordene los datos. ¿Qué conclusión puede deducir de estos datos?

- **2-15** La agencia de protección ambiental estadounidense tomó muestras de agua de 12 ríos y arroyos que desembocan en el lago Erie. Las muestras se probaron en los laboratorios de la agencia y clasificadas según la cantidad de contaminantes sólidos suspendidos en cada muestra. Los resultados de la prueba se dan en la tabla siguiente:

Muestra	1	2	3	4	5	6
Contaminantes (ppm)	37.2	51.7	68.4	54.2	49.9	33.4
Muestra	7	8	9	10	11	12
Contaminantes (ppm)	39.8	52.7	60.0	46.1	38.5	49.1

- Ordene los datos en un arreglo descendente.
 - Determine el número de muestras con un contenido de contaminantes entre 30.0 y 39.9, 40.0 y 49.9, 50.0 y 59.9 y entre 60.0 y 69.9.
 - Si 45.0 es el número que utiliza la agencia de protección ambiental para indicar una contaminación excesiva, ¿cuántas muestras serán clasificadas como excesivamente contaminadas?
 - ¿Cuál es la distancia más grande entre dos muestras consecutivas cualesquiera?
- **2-16** Suponga que el personal de admisiones al que nos referimos al analizar la tabla 2-1 de la página 12 desea examinar la relación entre la diferencial de un estudiante que realiza el examen de admisión (la diferencia entre el resultado real del examen y el esperado según el promedio general del nivel anterior) y la dispersión de los promedios generales del estudiante en el bachillerato y la universidad (la diferencia entre el promedio general de la universidad y el del nivel anterior). El personal de admisiones utiliza los datos siguientes:

Prom. bach.	Prom. univ.	Resultado examen	Prom. bach.	Prom. univ.	Resultado examen
3.6	2.5	1,100	3.4	3.6	1,180
2.6	2.7	940	2.9	3.0	1,010
2.7	2.2	950	3.9	4.0	1,330
3.7	3.2	1,160	3.2	3.5	1,150
4.0	3.8	1,340	2.1	2.5	940
3.5	3.6	1,180	2.2	2.8	960
3.5	3.8	1,250	3.4	3.4	1,170
2.2	3.5	1,040	3.6	3.0	1,100
3.9	3.7	1,310	2.6	1.9	860
4.0	3.9	1,330	2.4	3.2	1,070

Además, el personal de admisiones ha recibido la siguiente información del servicio de pruebas educativas:

Prom. bach.	Resultado examen	Prom. bach.	Resultado examen
4.0	1,340	2.9	1,020
3.9	1,310	2.8	1,000
3.8	1,280	2.7	980
3.7	1,250	2.6	960

Continúa

Prom. bach.	Resultado examen	Prom. bach.	Resultado examen
3.6	1,220	2.5	940
3.5	1,190	2.4	920
3.4	1,160	2.3	910
3.3	1,130	2.2	900
3.2	1,100	2.1	880
3.1	1,070	2.0	860
3.0	1,040		

- Ordene estos datos en un arreglo descendente de las dispersiones. (Considere positivo un aumento en el promedio general de universidad respecto al de bachillerato y como negativo el caso contrario.) Incluya con cada dispersión la diferencial de admisión correspondiente. (Considere negativo un resultado de admisión que esté por debajo del esperado y positivo uno que esté por arriba.)
- ¿Cuál es la dispersión más común?
- Para la dispersión del inciso b), ¿cuál es la diferencial de admisión más común?
- ¿A qué conclusiones llega, partiendo del análisis que ha hecho?

Soluciones a los ejercicios de autoevaluación

EA 2-1

7 intervalos		13 intervalos			
Clase	Frecuencia relativa	Clase	Frecuencia relativa	Clase	Frecuencia relativa
30-39	0.02	35-39	0.02	70-74	0.10
40-49	0.06	40-44	0.04	75-79	0.10
50-59	0.16	45-49	0.02	80-84	0.12
60-69	0.32	50-54	0.08	85-89	0.04
70-79	0.20	55-59	0.08	90-94	0.04
80-89	0.16	60-64	0.10	95-99	0.04
90-99	0.08	65-69	0.22		1.00
	1.00				

- Como se puede saber a partir de cualquiera de las distribuciones, cerca de 90% de los participantes tiene más de 50 años, por lo que el programa no cumple la política.
- En este caso, es igualmente sencillo usar las dos.
- La distribución de 13 intervalos da una mejor estimación porque tiene una clase de 45-49, mientras que la distribución de 7 intervalos agrupa todas las observaciones entre 40 y 49.

EA 2-2 Ordenamiento de datos según el promedio general de bachillerato:

Prom. bach.	Prom. univ.	Prom. bach.	Prom. univ.
4.0	3.9	3.4	3.4
4.0	3.8	3.2	3.5
3.9	4.0	2.9	3.0
3.9	3.7	2.7	2.2
3.7	3.2	2.6	2.7
3.6	3.0	2.6	1.9
3.6	2.5	2.4	3.2
3.5	3.8	2.2	3.5
3.5	3.6	2.2	2.8
3.4	3.6	2.1	2.5

Ordenamiento de datos según el promedio general en la universidad:

Prom. bach.	Prom. univ.	Prom. bach.	Prom. univ.
4.0	3.9	3.2	3.7
3.9	4.0	3.2	2.4

Continúa

Prom. bach.	Prom. univ.	Prom. bach.	Prom. univ.
3.8	4.0	3.0	3.6
3.8	3.5	3.0	2.9
3.7	3.9	2.8	2.2
3.6	3.5	2.7	2.6
3.6	3.4	2.5	3.6
3.5	3.2	2.5	2.1
3.5	2.2	2.2	2.7
3.4	3.4	1.9	2.6

De estos ordenamientos se puede ver que los promedios generales altos en un nivel tienden a relacionarse con promedios generales altos en el otro, aunque hay algunas excepciones.

2.4 Construcción de una distribución de frecuencias

Clasifique los datos

Ahora que ya hemos aprendido a dividir una muestra en clases, podemos tomar los datos sin procesar y construir una distribución de frecuencias. Para resolver el problema de los telares para alfombra de la primera página del capítulo, siga estos tres pasos:

Divida el rango total de los datos en clases de igual tamaño

1. Decida el tipo y número de clases para dividir los datos. En este caso, ya hemos escogido clasificar los datos de acuerdo con la medida cuantitativa del número de yardas producidas, más que con respecto a un atributo cualitativo como color o estampado. En seguida, necesitamos decidir cuántas clases distintas usar y el alcance que cada clase debe cubrir. El rango total debe dividirse entre clases *iguales*; esto es, el ancho del intervalo, tomado desde el principio de una clase hasta el principio de la siguiente, debe ser el mismo para todas. Si escogemos un ancho de 0.5 yardas para cada clase de la distribución, las clases serán las que se muestran en la tabla 2-12.

Problemas con clases desiguales

Si las clases fueran desiguales y el ancho de los intervalos variara de una clase a otra, tendríamos una distribución mucho más difícil de interpretar que una con intervalos iguales. ¡Imagine lo difícil que sería interpretar los datos de la tabla 2-13!

Use de 6 a 15 clases

El número de clases depende del número de datos puntuales y del alcance de los datos recolectados. Cuantos más datos puntuales se tengan o más grande sea el rango, más clases se necesitarán para dividir los datos. Desde luego, si solamente tenemos 10 datos puntuales, no tendría sentido plantear 10 clases. Como regla general, los estadísticos rara vez utilizan menos de seis y más de 15 clases.

Determine el ancho de los intervalos de clase

Debido a que necesitamos hacer los intervalos de clase de igual tamaño, el número de clases determina el ancho de cada clase. Para hallar el ancho de los intervalos podemos utilizar esta ecuación:

Ancho de un intervalo de clase	
Ancho de los intervalos de clase =	$\frac{\text{Valor unitario siguiente después del valor más grande de los datos} - \text{Valor más pequeño de los datos}}{\text{Número total de intervalos de clase}} \quad [2-1]$

Tabla 2-12	Clase en yardas	Frecuencia
Producción diaria de una muestra de 30 telares para alfombras con intervalos de clase de 0.5 yardas	15.1-15.5	2
	15.6-16.0	16
	16.1-16.5	8
	16.6-17.0	4
		<u>30</u>

Tabla 2-13	Clase	Ancho de intervalos de clase	Frecuencia
Producción diaria de una muestra de 30 telares para alfombra utilizando intervalos de clase desiguales	15.1-15.5	15.6 – 15.1 = 0.5	2
	15.6-15.8	15.9 – 15.6 = 0.3	8
	15.9-16.1	16.2 – 15.9 = 0.3	9
	16.2-16.5	16.6 – 16.2 = 0.4	7
	16.6-16.9	17.0 – 16.6 = 0.4	4
			<u>30</u>

Debemos utilizar el *siguiente valor de las mismas unidades*, ya que estamos midiendo el *intervalo* entre el primer valor de una clase y el primer valor de la siguiente. En nuestro estudio de los telares, el último valor es 16.9, de modo que el siguiente valor es 17.0. Como estamos utilizando seis clases en este ejemplo, el ancho de cada clase será:

$$\begin{aligned}
 & \frac{\text{Valor unitario siguiente después del valor más grande de los datos} - \text{Valor más pequeño de los datos}}{\text{Número total de intervalos de clase}} && [2-1] \\
 & = \frac{17.0 - 15.2}{6} \\
 & = \frac{1.8}{6} \\
 & = 0.3 \text{ yd} \leftarrow \text{ancho de los intervalos de clase}
 \end{aligned}$$

Examine los resultados

El paso 1 está completo. Hemos decidido clasificar los datos según las mediciones cuantitativas de cuántas yardas de alfombra fueron producidas. Definimos seis clases para cubrir el rango de 15.2 a 16.9 y, como resultado de ello, utilizamos 0.3 yardas como el ancho de nuestros intervalos de clase.

Construya las clases y cuente las frecuencias

2. Clasifique los datos puntuales en clases y cuente el número de observaciones que hay en cada una. Hicimos esto en la tabla 2-14. Cada dato puntual entra al menos en una clase y ningún dato puntual entra en más de una clase. En consecuencia, nuestras clases son completamente incluyentes y mutuamente excluyentes. Observe que el límite inferior de la primera clase corresponde al dato puntual menor de la muestra, y que el límite superior de la última clase corresponde al de dato puntual mayor.

3. Ilustre los datos en un diagrama. (Vea la figura 2-1.)

Estos tres pasos nos permiten organizar los datos en forma tanto tabular como gráfica. En este caso, nuestra información se muestra en la tabla 2-14 y en la figura 2-1. Estas dos distribuciones de frecuencias omiten algunos de los detalles contenidos en los datos sin procesar de la tabla 2-3, pero nos facilitan la observación de los patrones contenidos en ellos. Por ejemplo, una característica obvia es que la clase 15.8-16.0 contiene el mayor número de elementos; mientras que la 15.2-15.4, contiene el menor.

Tabla 2-14	Clase	Frecuencia
Producción diaria de una muestra de 30 telares para alfombra con intervalos de clase de 0.3 yardas	15.2-15.4	2
	15.5-15.7	5
	15.8-16.0	11
	16.1-16.3	6
	16.4-16.6	3
	16.7-16.9	<u>3</u>
	<u>30</u>	

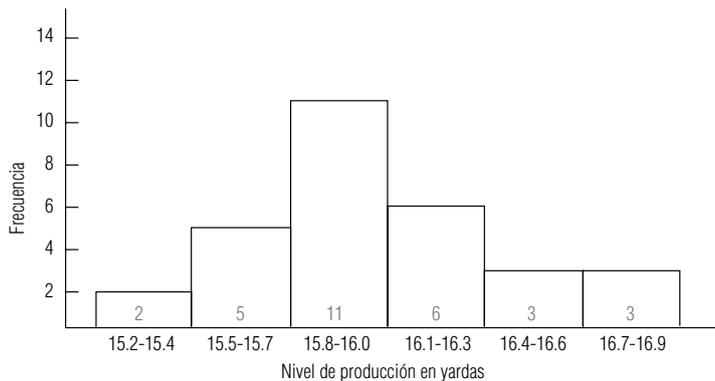


FIGURA 2-1

Distribución de frecuencia de los niveles de producción de una muestra de 30 telares para alfombra con intervalos de clase de 0.3 yardas

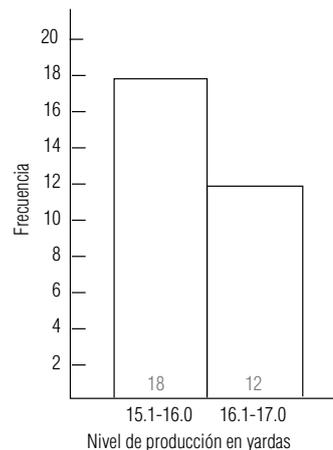


FIGURA 2-2

Distribución de frecuencias de los niveles de producción de una muestra de 30 telares para alfombra con intervalos de clase de una yarda

Detecte las tendencias

Observe, en la figura 2-1, que las frecuencias de las clases con ancho de 0.3 yardas siguen una secuencia regular: el número de datos puntuales empieza con dos para la primera clase, aumenta hasta cinco en la segunda, alcanza 11 en la tercera clase, disminuye a seis y luego cae a tres en la quinta y sexta clases. Tendremos que, cuanto más ancho sea el intervalo de clase, más suave será la progresión. Sin embargo, si las clases son demasiado anchas, podemos perder mucha información, al grado de que la gráfica carezca de significado. Por ejemplo, si compactamos la figura 2-1 y tomamos sólo dos categorías, oscurecemos el patrón. Esto se hace evidente en la figura 2-2.

Uso de la computadora para construir distribuciones de frecuencias

Los cálculos manuales son tediosos

A lo largo del presente texto, usaremos ejemplos sencillos para ilustrar cómo realizar muchos tipos de análisis estadísticos. Con tales ejemplos, usted puede aprender qué tipo de cálculos debe hacer. Esperamos que también sea capaz de entender los conceptos que hay detrás de los cálculos, de modo que pueda apreciar por qué algunos cálculos en particular son los apropiados. Sin embargo, prevalece la cuestión de hacer los cálculos de forma manual es tedioso, cansado y proclive al error. Además, muchos problemas reales tienen tal cantidad de datos que no es posible hacer los cálculos a mano.

Paquetes de software para análisis estadísticos

Por esta razón, casi todos los análisis estadísticos reales se hacen con ayuda de computadoras. Usted prepara los datos que introducirá a la máquina e interpreta los resultados y decide las acciones adecuadas, pero la computadora realiza todos los cálculos numéricos. Existen muchos paquetes de software ampliamente utilizados para hacer análisis estadísticos. Entre otros, están Minitab, SAS, Excel, SPSS y SYSTAT.* Pero emplearemos Minitab, SPS y Excel para ilustrar los resultados típicos que producen.

Uso de los datos de calificaciones

El apéndice 10 contiene los datos de calificaciones de los 199 estudiantes que usaron este texto durante 2002. En la figura 2-3 se usó Minitab para crear una distribución de frecuencias de las calificaciones totales sin procesar de los estudiantes del curso. Los valores de la columna TOTPOR10 son los puntos medios o marcas de clase. Con frecuencia también son de interés las *distribuciones de frecuencias bivariadas*, en las que los datos se clasifican de acuerdo con dos atributos diferentes. En la figura 2-4 se ilustra una distribución que muestra las calificaciones con letras en las seis sec-

*Minitab es marca registrada de Minitab, Inc., University Park, Pa. SAS es marca registrada de SAS Institute, Inc., Cary N. C. SPSS es marca registrada de SPSS, Inc., Chicago, Ill. SYSTAT es marca registrada de SYSTAT, Inc. Evanston, Ill. Excel es marca registrada de Microsoft.

FIGURA 2-3

Distribución de frecuencias de calificaciones totales sin procesar obtenida con Minitab

Resumen estadístico para variables discretas					
TOTPOR10	Cuenta	Porcentaje	Ctaacum	Pctacum	
25	1	0.50	1	0.50	
35	1	0.50	2	1.01	
45	9	4.52	11	5.53	
55	27	13.57	38	19.10	
65	68	34.17	106	53.27	
75	65	32.66	171	85.93	
85	26	13.07	197	98.99	
95	2	1.01	199	100.00	
N=	199				

Estadísticas tabuladas

RENGLONES: CALIFNUM

COLUMNAS: SECCIÓN

	1	2	3	4	5	6	TODO
0	2 1.01	3 1.51	0 --	1 0.50	3 1.51	2 1.01	11 5.53
1	3 1.51	6 3.02	5 2.51	2 1.01	4 2.01	6 3.02	26 13.07
2	2 1.01	2 1.01	1 0.50	2 1.01	7 3.52	4 2.01	18 9.05
3	9 4.52	11 5.53	3 1.51	9 4.52	6 3.02	6 3.02	44 22.11
4	3 1.51	6 3.02	10 5.03	6 3.02	7 3.52	2 1.01	34 17.09
5	1 0.50	5 2.51	5 2.51	1 0.50	0 --	3 1.51	15 7.54
6	2 1.01	5 2.51	3 1.51	2 1.01	2 1.01	3 1.51	17 8.54
7	1 0.50	1 0.50	1 0.50	2 1.01	1 0.50	1 0.50	7 3.52
8	2 1.01	2 1.01	8 4.02	1 0.50	3 1.51	0 --	16 8.04
9	2 1.01	5 2.51	1 0.50	0 --	3 1.51	0 --	11 5.53
TODO	27 13.57	46 23.12	37 18.59	26 13.07	36 18.09	27 13.57	199 100.00

FIGURA 2-4

Distribución de frecuencias bivariada obtenida con Minitab que muestra las calificaciones en cada sección

CONTENIDO DE CELDAS --

CUENTA
% DE TABLA

ciones de la clase. La variable CALIFNUM tiene valores de 0 a 9, que corresponden a las calificaciones con letras F, D, C-, C, C+, B-, B, B+, A- y A.

El apéndice 11 contiene los datos correspondientes a las ganancias del último trimestre de 1989 de 224 compañías, los cuales fueron publicados en *The Wall Street Journal* durante la segunda semana de febrero de 1990. La figura 2-5 ilustra la distribución de frecuencias obtenida con Minitab de las ganancias del último trimestre de 1989. La variable Q489 representa las ganancias del último trimestre de 1989, en cifras redondeadas a dólares.

Debido a que las compañías incluidas en la New York Stock Exchange (Bolsa de Valores de Nueva York) (3) tienden a tener características financieras diferentes a las enumeradas en la American Stock Exchange (Bolsa de Valores Estadounidense) (2), y debido también a que éstas, a su vez, son

Resumen estadístico para variables discretas

Q489	Cuenta	Porcentaje	Ctaacum	Pctacum
-5	1	0.45	1	0.45
-4	2	0.89	3	1.34
-2	1	0.45	4	1.79
-1	9	4.02	13	5.80
0	164	73.21	177	79.02
1	43	19.20	220	98.21
2	2	0.89	222	99.11
5	2	0.89	224	100.00
N=	224			

FIGURA 2-5

Distribución de frecuencias obtenida con Minitab de las calificaciones del último trimestre de 1989

distintas a las de las compañías enumeradas en “al contado” (1), también utilizamos MYSTAT para producir la distribución bivariada de los mismos datos en la figura 2-6.

Estadísticas tabuladas

RENGLONES: Q489 COLUMNAS: TRANSACCIÓN

	1	2	3	TODO
-5	0	0	1	1
	--	--	100.00	100.00
	--	--	1.33	0.45
	--	--	0.45	0.45
-4	1	0	1	2
	50.00	--	50.00	100.00
	0.90	--	1.33	0.89
	0.45	--	0.45	0.89
-2	1	0	0	1
	100.00	--	--	100.00
	0.90	--	--	0.45
	0.45	--	--	0.45
-1	5	2	2	9
	55.56	22.22	22.22	100.00
	4.50	5.26	2.67	4.02
	2.23	0.89	0.89	4.02
0	97	31	36	164
	59.15	18.90	21.95	100.00
	87.39	81.58	48.00	73.21
	43.30	13.84	16.07	73.21
1	7	4	32	43
	16.28	9.30	74.42	100.00
	6.31	10.53	42.67	19.20
	3.12	1.79	14.29	19.20
2	0	0	2	2
	--	--	100.00	100.00
	--	--	2.67	0.89
	--	--	0.89	0.89
5	0	1	1	2
	--	50.00	50.00	100.00
	--	2.63	1.33	0.89
	--	0.45	0.45	0.89
TODO	111	38	75	224
	49.55	16.96	33.48	100.00
	100.00	100.00	100.00	100.00
	49.55	16.96	33.48	100.00

FIGURA 2-6

Distribución de frecuencias bivariadas obtenida con Minitab en donde se muestran los ingresos de cada transacción

CONTENIDO DE LAS CELDAS --

CUENTA
% DE RENGLÓN
% DE COLUMNA
% DE TABLA

SUGERENCIAS**Y****SUPOSICIONES**

Es necesario ser extremadamente cuidadosos al elegir las clases cuando se elabora una distribución de frecuencias, incluso cuando un programa de computación define las clases. Por ejemplo, un programa de cómputo puede clasificar las respuestas a un estudio de mercado de acuerdo con la edad del consumidor: 15-19, 20-24, 25-29,

y así sucesivamente. Pero si el producto bajo estudio está dirigido a estudiantes universitarios, tendría más sentido agrupar los datos en las categorías: 18, 19-22 y 23 o mayores. No pierda de vista que el uso de una computadora para efectuar cálculos estadísticos no sustituye al sentido común.

Ejercicios 2-4

Ejercicios de autoevaluación

- EA 2-3** La compañía High Performance Bicycle Products de Chapel Hill, Carolina del Norte, hizo un muestreo de sus registros de embarque para cierto día con los siguientes resultados:

Tiempo entre la recepción de una orden y su entrega (en días)

4	12	8	14	11	6	7	13	13	11
11	20	5	19	10	15	24	7	29	6

Construya una distribución de frecuencias para estos datos y una distribución de frecuencias relativas. Use intervalos de 6 días.

- ¿Qué puede asegurar acerca de la efectividad del procesamiento de pedidos a partir de la distribución de frecuencias?
 - Si la compañía desea asegurar que la mitad de sus entregas se hagan en 10 días o menos, ¿puede determinar, a partir de la distribución de frecuencias, si han logrado esta meta?
 - ¿Qué puede hacer con los datos si tiene una distribución de frecuencias relativas, que es difícil lograr con sólo una distribución de frecuencias?
- EA 2-4** El señor Franks, un ingeniero de seguridad de Mars Point Nuclear Power Generating Station, elaboró una tabla de la temperatura pico del reactor cada día durante el último año y preparó la siguiente distribución de frecuencias:

Temperaturas en °C	Frecuencia
Menos de 500	4
501-510	7
511-520	32
521-530	59
530-540	82
550-560	65
561-570	33
571-580	28
580-590	27
591-600	23
Total	360

Enumere y explique los errores que pueda encontrar en la distribución del señor Franks.

Aplicaciones

- 2-17** La Universal Burger está preocupada por la cantidad de desperdicio que genera, por lo que obtuvo una muestra del desperdicio de hamburguesas desde el año pasado con los siguientes resultados:

Número de hamburguesas descartadas durante un turno

2	16	4	12	19	29	24	7	19
22	14	8	24	31	18	20	16	6

Construya una distribución de frecuencias y una distribución de frecuencias relativas para estos datos. Utilice intervalos de 5 hamburguesas.

- Una de las metas de Universal Burger es que al menos 75% de los turnos tengan no más de 16 hamburguesas desperdiciadas. ¿Puede determinar a partir de la distribución de frecuencia si se logra esta meta?
 - ¿Qué porcentaje de turnos tienen desperdicios de 21 hamburguesas o menos? ¿Qué distribución usó para determinar su respuesta?
- **2-18** Remítase a la tabla 2-2 y construya una distribución de frecuencias relativas usando intervalos de 4.0 lb/pulg². ¿Qué se puede concluir a partir de esta distribución?
- **2-19** La Oficina de Estadística Laboral ha definido una muestra de 30 comunidades de todo Estados Unidos y compiló los precios de productos básicos en cada comunidad al principio y al final del mes de agosto, con el fin de encontrar cuánto ha variado aproximadamente el Índice de Precios al Consumidor (IPC) (CPI, Consumer Price Index) durante ese mes. Los cambios porcentuales en los precios para las 30 comunidades son:

0.7	0.4	-0.3	0.2	-0.1	0.1	0.3	0.7	0.0	-0.4
0.1	0.5	0.2	0.3	1.0	-0.3	0.0	0.2	0.5	0.1
-0.5	-0.3	0.1	0.5	0.4	0.0	0.2	0.3	0.5	0.4

- Ordene los datos en un arreglo ascendente.
 - Utilizando las siguientes cuatro clases de igual tamaño, construya una distribución de frecuencias: -0.5 a -0.2, -0.1 a 0.2, 0.3 a 0.6 y 0.7 a 1.0.
 - ¿Cuántas comunidades tienen precios que no han cambiado o que se han incrementado en menos de 1.0%?
 - ¿Son estos datos continuos o discretos?
- **2-20** Sarah Anne Ralpp, presidenta de Baggit, Inc., acaba de obtener algunos datos sin procesar de una investigación de mercado que su compañía hizo recientemente. La investigación se llevó a cabo para determinar la efectividad del nuevo lema publicitario de la empresa, “Cuando termine con todo, ¡embólselo!” Para determinar el efecto del lema sobre las ventas de los empaques Embólselo para comida, se interrogó a 20 personas respecto a cuántos empaques compraba mensualmente antes de conocer el lema publicitario y cuántos compra después que éste se utilizó en una campaña. Los resultados fueron los siguientes:

Antes/Después		Antes/Después		Antes/Después		Antes/Después	
4	3	2	1	5	6	8	10
4	6	6	9	2	7	1	3
1	5	6	7	6	8	4	3
3	7	5	8	8	4	5	7
5	5	3	6	3	5	2	2

- Construya las distribuciones de frecuencias simples y relativas para las respuestas dadas “antes”, utilizando como clases los intervalos 1 a 2, 3 a 4, 5 a 6, 7 a 8 y 9 a 10.
 - Haga lo mismo que en el inciso anterior para los datos de “después”.
 - Mencione la razón fundamental por la cual tiene sentido utilizar las mismas clases para ambos tipos de respuestas, “antes” y “después”.
 - Para cada pareja de respuestas “antes/después”, reste la respuesta “antes” de la respuesta “después” para obtener el número que llamaremos “cambio” (ejemplo: 3 - 4 = -1), y construya ambas distribuciones de frecuencias, simples y relativas, para la categoría “cambio”, utilizando las clases -5 a -4, -3 a -2, -1 a 0, 1 a 2, 3 a 4 y 5 a 6.
 - Con base en su análisis, establezca si el nuevo lema publicitario ha contribuido a las ventas, e indique una o dos razones que sustenten su conclusión.
- **2-21** A continuación presentamos las edades de 30 personas que compraron grabadoras de video en la tienda de música Symphony durante la semana pasada:

26	37	40	18	14	45	32	68	31	37
20	32	15	27	46	44	62	58	30	42
22	26	44	41	34	55	50	63	29	22

- a) A partir de una observación de los datos como se presentan, ¿a qué conclusiones puede llegar rápidamente acerca del mercado de Symphony?
- b) Construya una clasificación cerrada de seis categorías. ¿Esta clasificación le permite llegar a alguna conclusión adicional sobre el mercado de Symphony?
- **2-22** Utilice los datos del ejercicio 2-21.
- a) Construya una clasificación de extremo abierto de cinco categorías. ¿Esta clasificación le permite llegar a alguna otra conclusión acerca del mercado de Symphony?
- b) Construya ahora una distribución de frecuencias relativas que corresponda a la clasificación de extremo abierto de cinco categorías. ¿Esta distribución le proporciona a Symphony información adicional útil acerca de su mercado? ¿Por qué?
- **2-23** John Lyon, dueño de la tienda de alimentos Fowler's de Chapel Hill, Carolina del Norte, tiene organizadas las compras que le hicieron sus clientes durante la semana anterior en la siguiente distribución de frecuencias:

Cant. gastada	Frecuencia
0.00- 0.99	50
1.00- 3.99	240
4.00- 6.99	300
7.00- 9.99	460
10.00-12.99	900
13.00-15.99	1,050
16.00-18.99	1,150
19.00-21.99	980
22.00-24.99	830
25.00-27.99	780
28.00-30.99	760
31.00-33.99	720
34.00-36.99	610
37.00-39.99	420
40.00-42.99	280
43.00-45.99	100
46.00-48.99	90

John dice que es incómodo tener 17 intervalos, cada uno definido por dos números. ¿Puede usted sugerir cómo simplificar la clasificación de los datos sin perder demasiado su valor?

- **2-24** A continuación puede ver los puntos medios de los intervalos de una distribución de frecuencias que representa el tiempo en minutos que tardaron los miembros del equipo de atletismo de una universidad en recorrer ocho kilómetros a campo traviesa.

25 35 45

- a) ¿Diría que el entrenador puede obtener información suficiente a partir de estos puntos medios para ayudar al equipo?
- b) Si la respuesta al inciso anterior es "no", ¿cuántos intervalos parecerían los apropiados?
- **2-25** Barney Mason examinó la cantidad de desperdicio de papas fritas (en libras) durante los últimos 6 meses en Universal Burger y creó la siguiente distribución de frecuencias:

Desperdicio de papas fritas en libras	Frecuencia
0.0- 3.9	37
4.0- 7.9	46
8.0-11.9	23
12.0-16.9	27
17.0-25.9	7
26.0-40.9	0
	<u>180</u>

Enumere y explique los errores que pueda encontrar en la distribución de Barney.

- **2-26** Construya una clasificación discreta y cerrada para las posibles respuestas a la pregunta acerca del “estado civil” en una solicitud de empleo. Construya también una clasificación discreta y de extremo abierto de tres categorías para las mismas respuestas.
- **2-27** Las listas de la bolsa de valores por lo general contienen el nombre de la compañía, las ofertas superior e inferior, el precio de cierre y el cambio con respecto al precio de cierre del día anterior. A continuación presentamos un ejemplo:

Nombre	Oferta superior	Oferta inferior	Cierre	Cambio
Sistemas asociados	$11\frac{1}{2}$	$10\frac{7}{8}$	$11\frac{1}{4}$	$+\frac{1}{2}$

¿Se trata de una distribución: a) de todos los valores de la Bolsa de Valores de Nueva York por industria, b) por precios al cierre de un día determinado y c) por cambios en los precios de un día dado?

¿La distribución es

- 1) cuantitativa o cualitativa?
- 2) continua o discreta?
- 3) de *extremo* abierto o cerrado?

¿Respondería de manera diferente al inciso c) si el cambio fuera expresado sólo como “mayor”, “menor” o “sin cambio”?

- **2-28** Los niveles de ruido en decibeles de un avión que despegar del aeropuerto del condado de Westchester fueron redondeados al decibel más cercano y agrupados en una distribución de frecuencias con intervalos que tienen puntos medios entre 100 y 130. Por debajo de los 100 decibeles no se consideran altos, mientras que cualquier nivel por arriba de los 140 decibeles resulta casi ensordecedor. Los miembros de la organización “Residentes de un barrio más tranquilo” están recabando datos para su pleito contra el aeropuerto, ¿esta distribución es adecuada para sus propósitos?
- **2-29** Utilice los datos del ejercicio 2-28. Si el abogado defensor del aeropuerto está recogiendo datos en su preparación para irse a juicio, ¿tomará los datos del punto medio (o marca de clase) de los intervalos del ejercicio 2-28 como favorables para sus objetivos?
- **2-30** El presidente de Ocean Airlines intenta hacer una estimación de cuánto se tardará el Departamento de Aeronáutica Civil (DAC) en decidir acerca de la solicitud de la compañía sobre una nueva ruta entre Charlotte y Nashville. Los asesores del presidente han organizado los siguientes tiempos de espera de las solicitudes formuladas durante el año anterior. Los datos se expresan en días, desde la fecha de la solicitud hasta la respuesta del DAC.

34	40	23	28	31	40	25	33	47	32
44	34	38	31	33	42	26	35	27	31
29	40	31	30	34	31	38	35	37	33
24	44	37	39	32	36	34	36	41	39
29	22	28	44	51	31	44	28	47	31

- a) Construya una distribución de frecuencias utilizando diez intervalos cerrados, igualmente espaciados. ¿Qué intervalo contiene el mayor número de datos puntuales?
 - b) Construya una distribución de frecuencias utilizando cinco intervalos cerrados, igualmente espaciados. ¿Qué intervalo contiene el mayor número de datos puntuales?
 - c) Si el presidente de la Ocean Airlines tiene una distribución de frecuencias relativas, ya sea para a) o para b), ¿le ayudará ésta para estimar la respuesta que necesita?
- **2-31** Con el propósito de hacer una evaluación de desempeño y un ajuste de cuotas, Ralph Williams estuvo inspeccionando las ventas de automóviles de sus 40 vendedores. En un periodo de un mes, tuvieron las siguientes ventas de automóviles:

7	8	5	10	9	10	5	12	8	6
10	11	6	5	10	11	10	5	9	13
8	12	8	8	10	15	7	6	8	8
5	6	9	7	14	8	7	5	5	14

- a) Con base en la frecuencia, ¿cuáles serían las marcas de clase (puntos medios de los intervalos) deseadas?
- b) Construya distribuciones de frecuencias y de frecuencias relativas con el mayor número posible de marcas de clase. Haga los intervalos igualmente espaciados y con un ancho de al menos dos automóviles.

- c) Si las ventas menores de siete automóviles por mes se consideran como un desempeño inaceptable, ¿cuál de los dos incisos, a) o b), es de mayor utilidad para identificar al grupo insatisfactorio de vendedores?
- **2-32** Kessler's Ice Cream Delight intenta mantener los 55 sabores diferentes de helado en existencia en cada una de sus sucursales. El director de investigación de mercado de la compañía sugiere que mantener un mejor registro para cada tienda es la clave para evitar quedarse sin existencias. Don Martin, director de operaciones, recolecta datos redondeados al medio galón más cercano sobre la cantidad diaria de helado de cada sabor que se vende. Nunca se venden más de 20 galones de un solo sabor al día.
- a) ¿La clasificación con respecto a los sabores es discreta o continua? ¿Abierta o cerrada?
 b) ¿La clasificación con respecto a la "cantidad de helado" es discreta o continua? ¿Abierta o cerrada?
 c) ¿Son los datos cuantitativos o cualitativos?
 d) ¿Qué le sugeriría a Don Martin para generar mejores datos para llevar a cabo una investigación de mercado?
- **2-33** Doug Atkinson es dueño y recogedor de boletos de un trasbordador que transporta pasajeros y automóviles desde Long Island hasta Connecticut. Doug tiene datos que indican el número de personas y de automóviles que han utilizado el trasbordador durante los dos últimos meses. Por ejemplo,

3 de julio Número de pasajeros 173 Número de automóviles 32

podría ser la información típica registrada durante un día. Doug ha establecido seis clases igualmente espaciadas para registrar el número diario de pasajeros y las marcas de clase son 84.5, 104.5, 124.5, 144.5, 164.5 y 184.5. Las seis clases igualmente espaciadas que construyó Doug para el número diario de automóviles tienen marcas de clase de 26.5, 34.5, 42.5, 50.5, 58.5 y 66.5. (Las marcas de clase son los puntos medios de los intervalos.)

- a) ¿Cuáles son los límites inferior y superior de las clases para el número de pasajeros?
 b) ¿Cuáles son los límites inferior y superior de las clases para el número de automóviles?

Soluciones a los ejercicios de autoevaluación

EA	2-3	Clase	1-6	7-12	13-18	19-24	25-30
		Frecuencia	4	8	4	3	1
		Frecuencia relativa	0.20	0.40	0.20	0.15	0.05

- a) Suponiendo que la compañía abre 6 días a la semana, se ve que 80% de las órdenes se entregan en 3 semanas o menos.
- b) Se puede decir que sólo entre 20 y 60% de las entregas se hacen en 10 días o menos, de manera que la distribución no genera suficiente información para determinar si la meta se cumple.
- c) Una distribución de frecuencias relativas permite presentar frecuencias como fracciones o porcentajes.
- EA **2-4** La distribución no contiene todos los datos. El valor 500°C no aparece, tampoco los puntos entre 541 y 549°C, inclusive. Además, la distribución está cerrada a la derecha, lo que elimina todos los datos mayores de 600°C. Estas omisiones pueden explicar el hecho de que el número total de observaciones sea sólo 360, en lugar de 365 como podría esperarse del conjunto de datos coleccionados durante un año. (Nota: no es absolutamente necesario que la distribución esté abierta a la derecha, en especial si no se registraron valores mayores que 600°C. Sin embargo, para que esté completa, la distribución debe ser continua en el intervalo seleccionado, aun cuando no haya datos en algunos intervalos.) Por último, las clasificaciones no son mutuamente excluyentes. Los puntos 530°C y 580°C están contenidos en más de un intervalo. Al crear un conjunto continuo de clasificaciones, debe tenerse cuidado de evitar este error.

2.5 Representación gráfica de distribuciones de frecuencias

Identificación de los ejes horizontal y vertical

Las figuras 2-1 y 2-2 son un avance de lo que estudiaremos a continuación: cómo presentar las distribuciones de frecuencias de manera gráfica. Las gráficas proporcionan datos en un diagrama de dos dimensiones. En el eje *horizontal* podemos mostrar los valores de la variable (la característica que estamos midiendo), como la producción de alfombras en yardas. En el eje *vertical* señalamos las frecuencias de las clases mostradas en el eje horizontal. De esta forma, la altura de las barras de la figura

Función de las gráficas

2-1 mide el número de observaciones que hay en cada clase señalada en el eje horizontal. Las gráficas de distribuciones de frecuencias y de distribuciones de frecuencias relativas son útiles debido a que resaltan y aclaran los patrones que no se pueden distinguir fácilmente en las tablas. Atraen la atención del que las observa hacia los patrones existentes en los datos. Las gráficas también ayudan a resolver problemas relacionados a las distribuciones de frecuencias; nos permiten estimar algunos valores con sólo una mirada y proporcionan una verificación visual sobre la precisión de nuestras soluciones.

Histogramas

Descripción de los histogramas

Las figuras 2-1 y 2-2 son dos ejemplos de histogramas. Un *histograma* consiste en una serie de rectángulos, cuyo ancho es proporcional al rango de los valores que se encuentran dentro de una clase, y cuya altura es proporcional al número de elementos que caen dentro de la clase. Si las clases empleadas en la distribución de frecuencias son del mismo ancho, entonces las barras verticales del histograma también tienen el mismo ancho. La altura de la barra correspondiente a cada clase representa el número de observaciones de la clase. Como consecuencia, el área contenida en cada rectángulo (base por altura) ocupa un porcentaje del área total de todos los rectángulos la cual es igual a la frecuencia absoluta de esa clase correspondiente respecto a todas las observaciones hechas.

Función de un histograma de frecuencias relativas

Un histograma que utiliza las frecuencias relativas de los datos puntuales de cada una de las clases, en lugar de usar el número real de puntos, se conoce como *histograma de frecuencias relativas*. Este tipo de histograma tiene la misma forma que un histograma de frecuencias absolutas construido a partir del mismo conjunto de datos. Esto es así debido a que en ambos, el tamaño relativo de cada rectángulo es la frecuencia de esa clase comparada con el número total de observaciones.

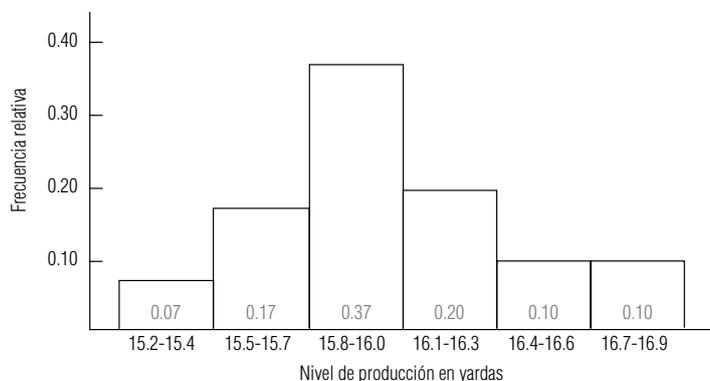
Recuerde que la frecuencia relativa de cualquier clase es el número de observaciones que entran en la clase, dividido entre el número total de observaciones hechas. La suma de todas las frecuencias relativas de cualquier conjunto de datos debe ser igual a 1.0. Con esto en mente, podemos convertir el histograma de la figura 2-1 en un histograma de frecuencias relativas como el presentado en la figura 2-7. Observe que la única diferencia entre éstos es el lado izquierdo de la escala del eje vertical. Mientras que la escala vertical del histograma de la figura 2-1 representa el número *absoluto* de observaciones de cada clase, la escala del histograma de la figura 2-7 es el número de observaciones de cada clase tomadas como una *fracción* del número total de observaciones.

Ventajas del histograma de frecuencias relativas

Poder presentar los datos en términos de la frecuencia relativa de las observaciones, más que en términos de la frecuencia absoluta, es de gran utilidad, ya que mientras los números absolutos pueden sufrir cambios (si probamos más telares, por ejemplo), la relación entre las clases permanece estable. El 20% de todos los telares puede entrar en la clase “16.1-16.3 yardas”, ya sea que probemos 30 o 300 telares. Resulta fácil comparar los datos de muestras de diferentes tamaños cuando utilizamos histogramas de frecuencias relativas.

FIGURA 2-7

Distribución de frecuencias relativas de los niveles de producción de una muestra de 30 telares para alfombra utilizando intervalos de clase de 0.3 yardas



Polígonos de frecuencias

Utilice los puntos medios en el eje horizontal

Aunque se utilizan menos, los polígonos de frecuencias son otra forma de representar gráficamente distribuciones tanto de frecuencias como de frecuencias relativas. Para construir un polígono de frecuencias señalamos éstas en el eje vertical y los valores de la variable que estamos midiendo en el eje horizontal, del mismo modo en que se hizo con el histograma. A continuación, graficamos cada frecuencia de clase trazando un punto sobre su punto medio y conectamos los puntos sucesivos resultantes con una línea recta para formar un polígono (una figura con muchos lados).

Añada dos clases

La figura 2-8 representa un polígono de frecuencias construido a partir de los datos de la tabla 2-14. Si compara esta figura con la 2-1, notará que se han agregado dos clases, una en *cada extremo* de la escala de valores observados. Éstas contienen cero observaciones, pero permiten que el polígono llegue al eje horizontal en ambos extremos de la distribución.

Conversión de un polígono de frecuencias en un histograma

¿De qué manera podemos convertir un polígono de frecuencias en un histograma? Un polígono de frecuencias es sólo una línea que conecta los puntos medios de todas las barras de un histograma. Por consiguiente, podemos reproducir el histograma mediante el trazado de líneas verticales desde los límites de clase (señalados en el eje horizontal) y, luego, conectando esas líneas con rectas horizontales a la altura de los puntos medios del polígono. En la figura 2-9 hicimos esto con líneas punteadas.

Construcción de un polígono de frecuencias relativas

Un polígono de frecuencias que utiliza frecuencias relativas de datos puntuales en cada una de las clases, en lugar del número real de puntos, se conoce como *polígono de frecuencias relativas*. Este polígono tiene la misma forma que el polígono de frecuencias construido a partir del mismo conjunto de datos, pero con una escala diferente en los valores del eje vertical. En lugar del número absoluto de observaciones, la escala representa el número de observaciones de cada clase expresadas como una fracción del total de observaciones.

Ventajas del histograma

Los histogramas y los polígonos de frecuencias son similares. ¿Por qué necesitamos ambos? Las ventajas de los histogramas son:

1. Los rectángulos muestran cada clase de la distribución por separado.
2. El área de cada rectángulo, en relación con el resto, muestra la proporción del número total de observaciones que se encuentran en esa clase.

FIGURA 2-8

Polígono de frecuencias del nivel de producción de una muestra de 30 telares para alfombra utilizando intervalos de clase de 0.3 yardas

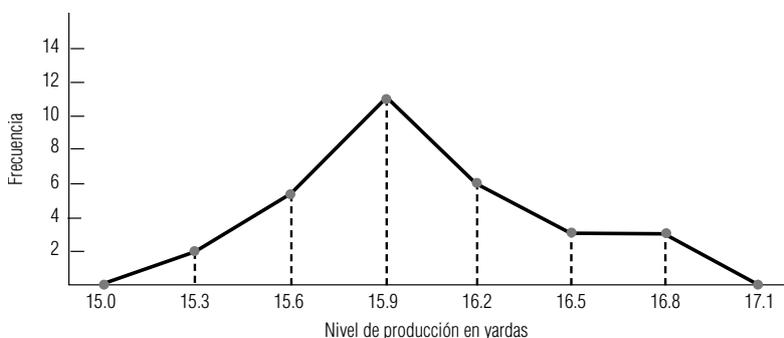
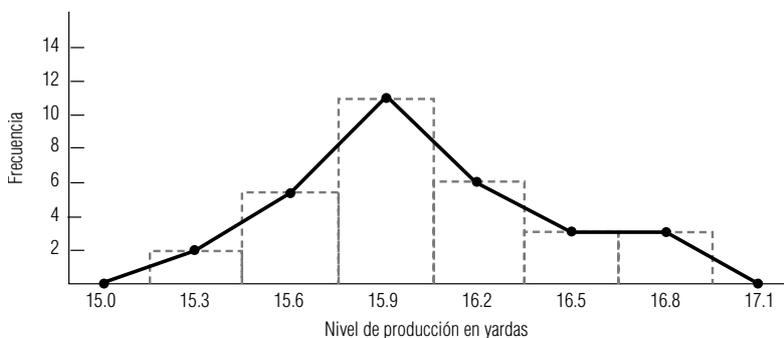


FIGURA 2-9

Histograma trazado a partir de los puntos del polígono de frecuencias de la figura 2-8



Ventajas de los polígonos

Los polígonos, por su parte, también poseen ciertas ventajas.

1. El polígono de frecuencias es más sencillo que su histograma correspondiente.
2. Bosqueja con más claridad un perfil del patrón de los datos.
3. El polígono se vuelve cada vez más suave y parecido a una curva conforme aumentamos el número de clases y el número de observaciones.

Creación de una curva de frecuencia

Un polígono como el que acabamos de describir, suavizado mediante el aumento de clases y de datos puntuales, se conoce como *curva de frecuencias*. En la figura 2-10, hemos utilizado el ejemplo de los telares para alfombra, pero en esta ocasión aumentamos el número de observaciones a 300 y el número de clases a 10. Note que conectamos los puntos con líneas curvas para tener una aproximación de la manera en que se vería el polígono si tuviéramos un gran número de datos puntuales e intervalos de clase muy pequeños.

Ojivas

Definición de distribución de frecuencias acumuladas

Una *distribución de frecuencias acumuladas* nos permite ver cuántas observaciones están por encima de ciertos valores, en lugar de hacer un mero registro del número de elementos que hay dentro de los intervalos. Por ejemplo, si deseamos saber cuántos telares tejen menos de 17.0 yardas, podemos utilizar una tabla que registre las frecuencias acumuladas “menores que” de nuestra muestra, como se presenta en la tabla 2-15.

Una ojiva “menor que”

La gráfica de una distribución de frecuencias acumuladas se conoce como *ojiva*. En la figura 2-11 se muestra la ojiva de la distribución de frecuencias de la tabla 2-15. Los puntos representados en la gráfica indican el número de telares que tienen una producción menor que el número de yardas mostrado en el eje horizontal. Note que el límite inferior de las clases de la tabla se convierte en el límite superior de la distribución acumulada de la ojiva.

Tabla 2-15	Clase	Frecuencia acumulada
Distribución de frecuencias “menores que” acumulada de los niveles de producción de una muestra de 30 telares para alfombra	Menos que 15.2	0
	Menos que 15.5	2
	Menos que 15.8	7
	Menos que 16.1	18
	Menos que 16.4	24
	Menos que 16.7	27
	Menos que 17.0	30

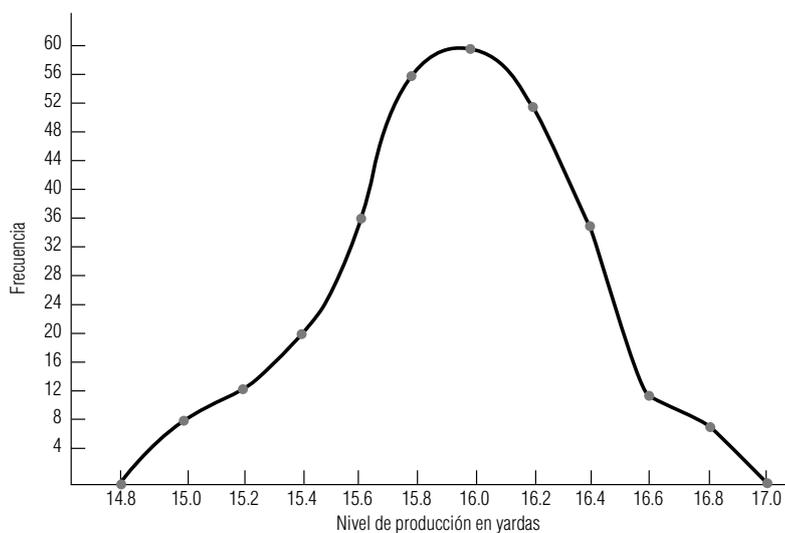
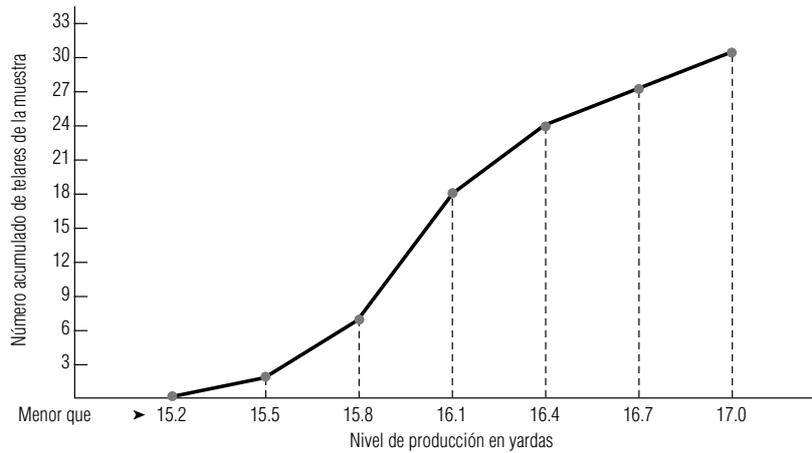


FIGURA 2-10

Curva de frecuencias de los niveles de producción de una muestra de 300 telares para alfombra utilizando intervalos de 0.2 yardas

FIGURA 2-11

Ojiva "menor que" de la distribución de niveles de producción de una muestra de 30 telares para alfombra



En algunas ocasiones, la información que utilizamos se presenta en términos de frecuencias acumuladas "mayores que". La ojiva adecuada para tal información tendría una inclinación hacia abajo y hacia la derecha, en lugar de tener una hacia arriba y a la derecha, como en la figura 2-11.

Ojivas de frecuencias relativas

Podemos construir una ojiva de una distribución de frecuencias relativas de la misma manera en que trazamos la ojiva de una distribución de frecuencias absolutas de la figura 2-11. Sólo habrá un cambio: la escala en el eje vertical. Como en la figura 2-7, esta escala debe señalar la *fracción* del número total de observaciones que caen dentro de cada clase.

Para construir una ojiva acumulada "menor que" en términos de frecuencias relativas, podemos remitirnos a una distribución de frecuencias relativas (como la de la figura 2-7) y construir una tabla utilizando los datos (como la tabla 2-16). Luego podemos convertir las cifras de dicha tabla en una ojiva (como en la figura 2-11). Observe que las figuras 2-11 y 2-12 son equivalentes y difieren sólo en la escala del eje vertical.

FIGURA 2-12

Ojiva "menor que" de la distribución de los niveles de producción de una muestra de 30 telares para alfombra utilizando frecuencias relativas

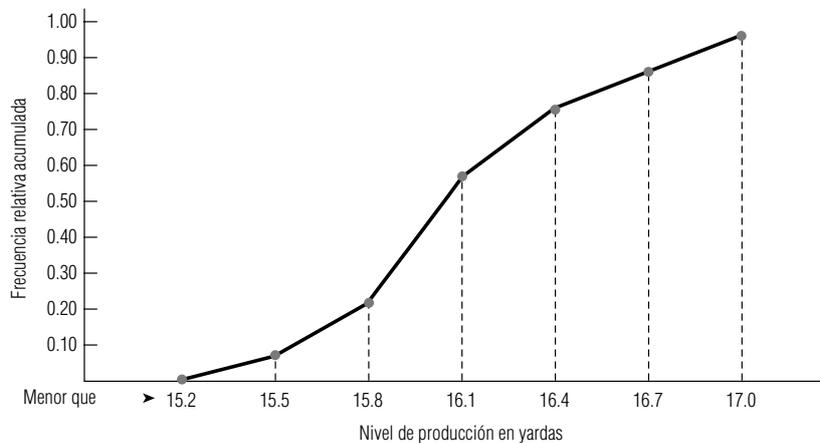
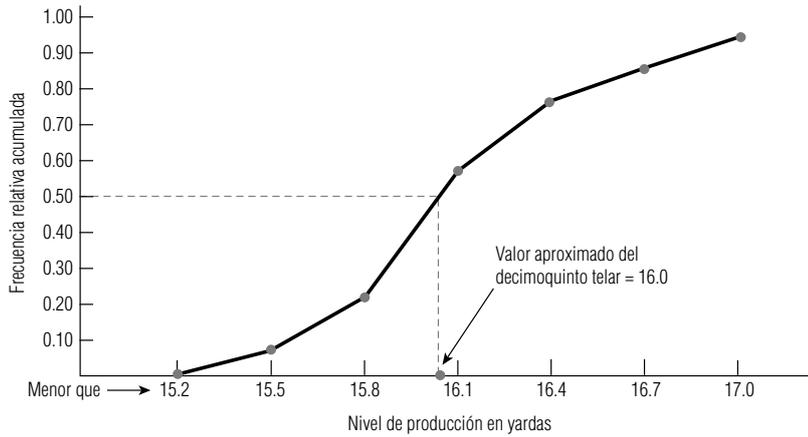


Tabla 2-16	Clase	Frecuencia acumulada	Frecuencia relativa acumulada
Distribución de frecuencias relativas acumuladas de los niveles de producción de una muestra de 30 telares para alfombra	Menor que 15.2	0	0.00
	Menor que 15.5	2	0.07
	Menor que 15.8	7	0.23
	Menor que 16.1	18	0.60
	Menor que 16.4	24	0.80
	Menor que 16.7	27	0.90
	Menor que 17.0	30	1.00

FIGURA 2-13

Ojiva "menor que" de la distribución de los niveles de producción de una muestra de 30 telares para alfombra, en donde se indica el valor medio aproximado del arreglo de datos original



Aproximación del arreglo de datos

Suponga que ahora trazamos una línea recta perpendicular al eje vertical en la marca 0.50 hasta intersectar a la ojiva (como en la figura 2-13). De esta manera, es posible leer un valor aproximado de 16.0 del nivel de producción del decimoquinto telar de un arreglo de 30.

Así pues, hemos regresado a la primera organización de datos estudiada en el presente capítulo. A partir del ordenamiento de datos podemos construir distribuciones de frecuencias; a partir de las distribuciones de frecuencias podemos construir distribuciones de frecuencias acumuladas; a su vez, a partir de éstas podemos trazar una ojiva. Y con base en esta ojiva podemos aproximar los valores provenientes del arreglo de datos. Sin embargo, en general no es posible recobrar los datos originales *exactos* a partir de ninguna de las representaciones gráficas analizadas.

Uso de la computadora para graficar distribuciones de frecuencias

Uso de SPSS para producir histogramas

Produzcamos algunos histogramas a partir de los datos de calificaciones presentados en el apéndice 10. La figura 2-14 proporciona un histograma de los promedios totales sin procesar de los estudiantes. Observe que éste tiene barras en posición horizontal en lugar de las barras verticales trazadas hasta ahora. Además, a la derecha de las barras, SPSS proporciona las frecuencias absolutas, las re-

Calificaciones de estadística en los negocios
Histograma

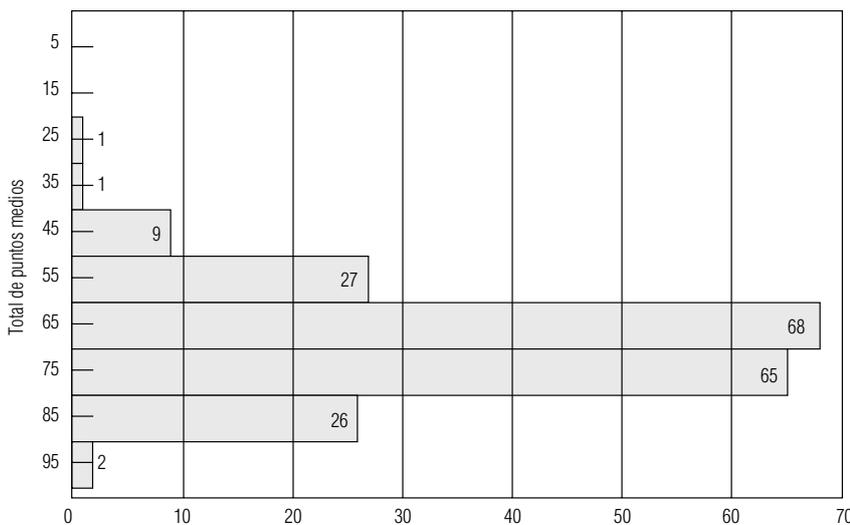
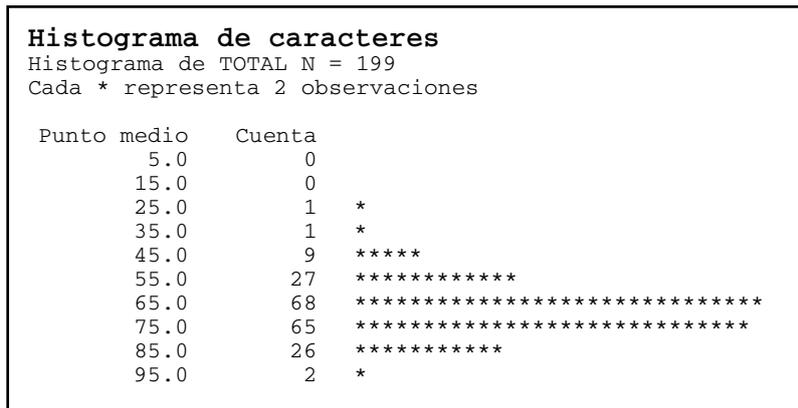


FIGURA 2-14

Histograma y distribución de frecuencias de promedios totales sin procesar, obtenidos con SPSS

FIGURA 2-15

Histograma y distribución de frecuencias para las calificaciones de Estadística para la administración, obtenidos con Minitab



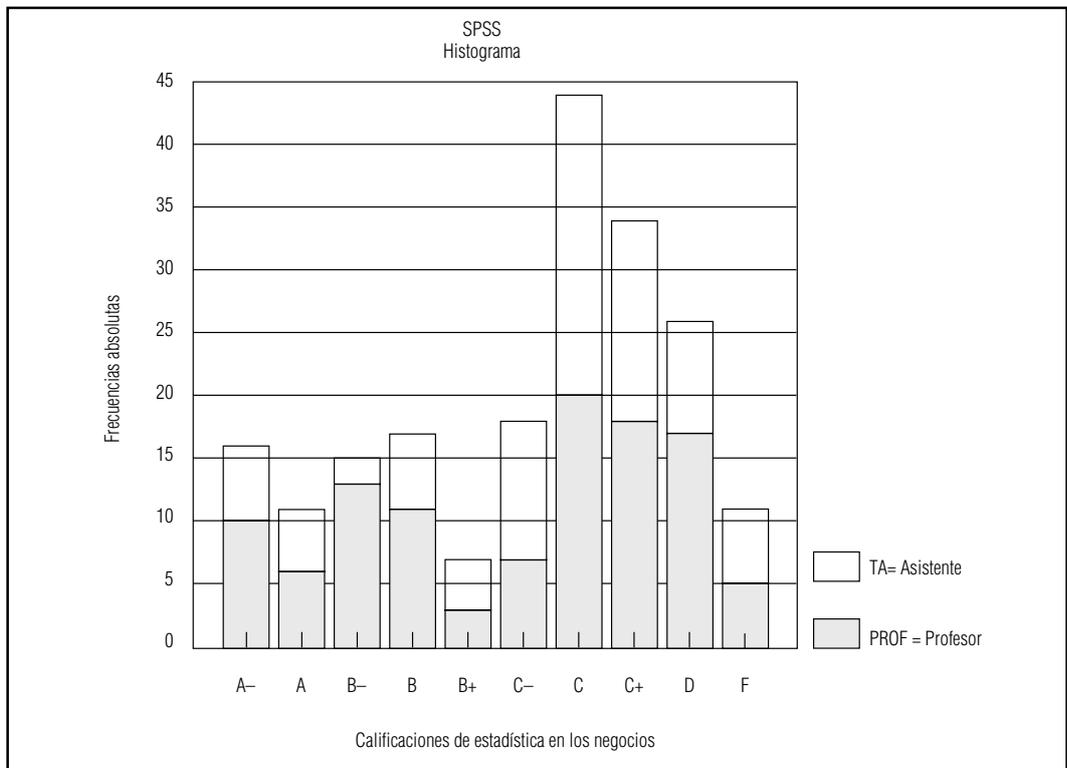
lativas y las acumuladas menores que (tanto absolutas como relativas). La figura 2-15 muestra la versión en Minitab.

En la figura 2-4 observamos una distribución de frecuencias bivariada. También podemos crear histogramas que contengan información acerca de dos variables. La figura 2-16 es un histograma SPSS vertical de las calificaciones con letras en el que cada barra está dividida en dos segmentos que muestran la fracción de estudiantes que obtuvieron esa calificación y que estaban en secciones impartidas por profesores y ayudantes de profesor (denotadas por P y T en las barras). La figura 2-17 es la versión de Minitab de las calificaciones, en la cual se emplean dos histogramas de CALIFNUM (vea la figura 2-4) dividida en INSTRNUM = 1, ayudantes de profesor, e INSTRUM = 2, profesor.

En la figura 2-18 se usó Minitab para producir un histograma de las ganancias del último trimestre de 1989 de las 224 compañías listadas en el apéndice 11. La figura 2-19 presenta histogramas de Minitab separados para las compañías 111 OTC, 38 ASE y 75 NYSE del conjunto de datos.

FIGURA 2-16

Histograma SPSS de calificaciones que muestra al tipo de instructor



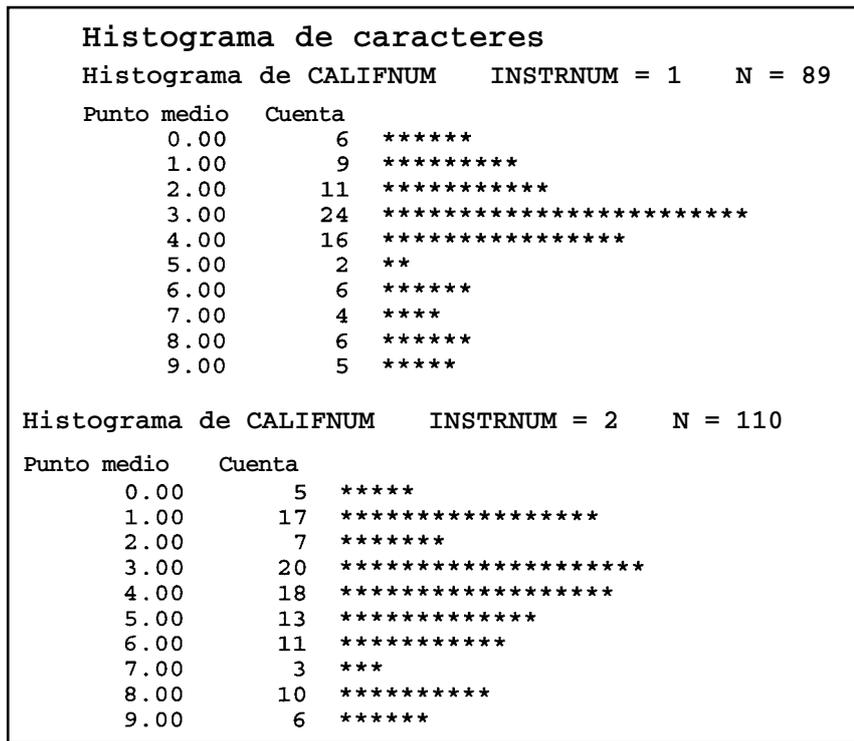


FIGURA 2-17

Versión Minitab de los histogramas de calificaciones

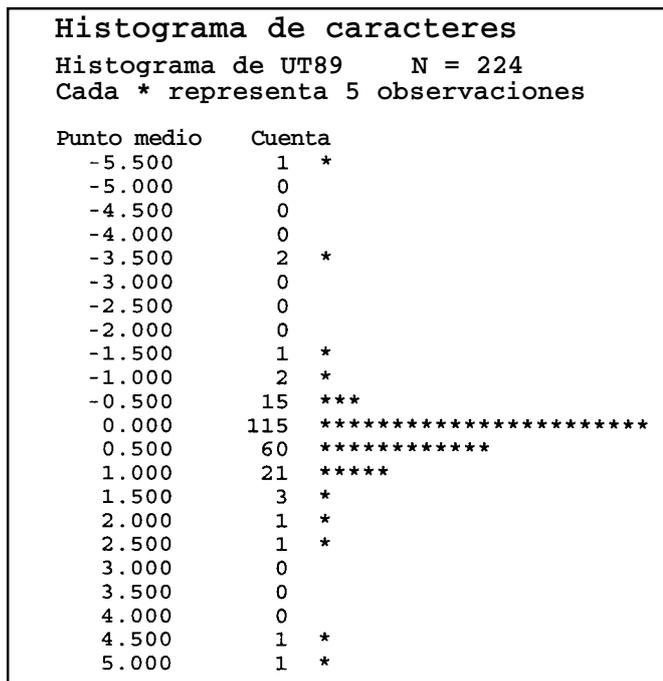


FIGURA 2-18

Histograma obtenido con Minitab para las ganancias del último trimestre de 1989

SUGERENCIAS Y SUPOSICIONES Quien haya dicho “una imagen vale mil palabras” entendía de manera intuitiva lo que se ha expuesto en esta sección. Al usar métodos gráficos para presentar datos se tiene una percepción rápida de los patrones y las tendencias, y de qué proporción de los datos está arriba o abajo de cierto valor. Advertencia: algunas publicaciones imprimen representaciones gráficas de datos (histogramas) de manera confusa usando un eje vertical que no llega a cero. Tenga cuidado con ellas pues las diferencias pequeñas pueden parecer grandes y el patrón que se observa puede llevar a conclusiones equivocadas.

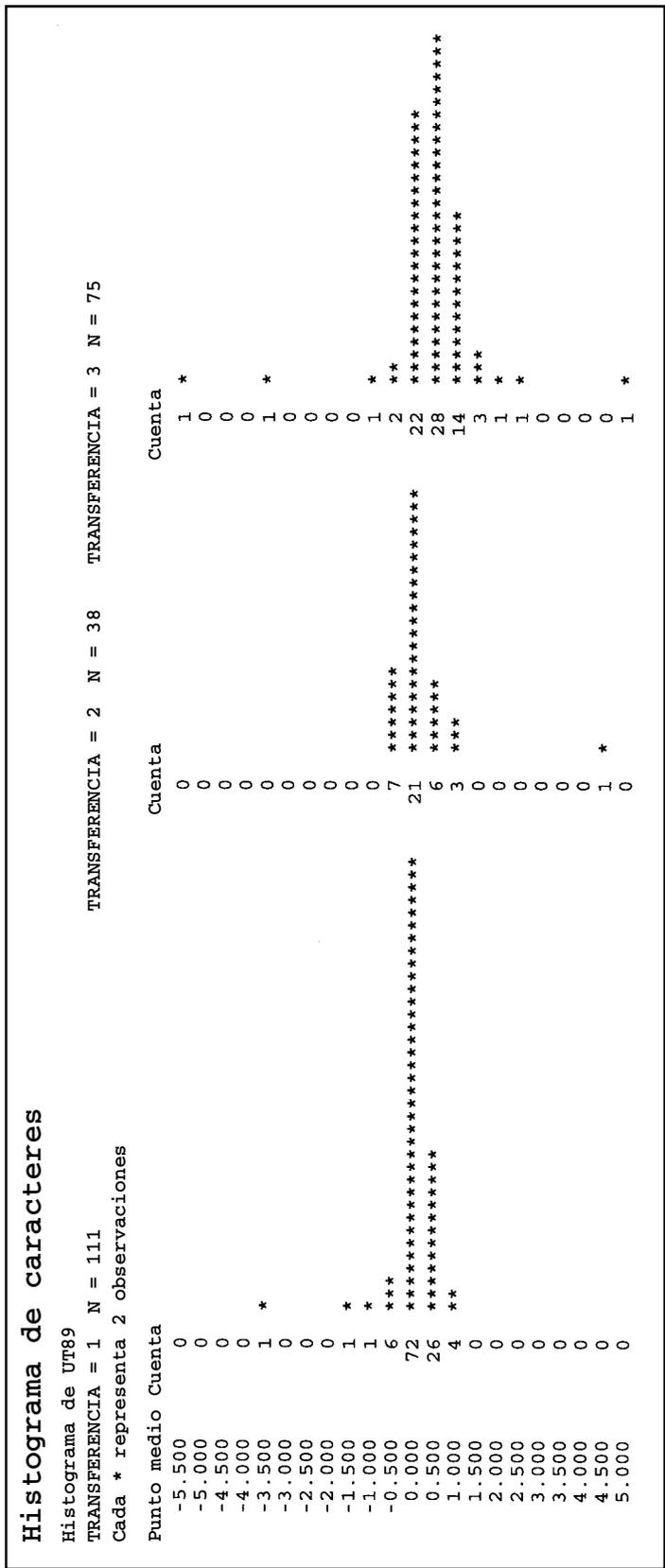


FIGURA 2-19 Histogramas Minitab separados por transacción para las ganancias del último trimestre de 1989

Ejercicios 2.5

Ejercicios de autoevaluación

- EA 2-5** Se presenta una distribución de frecuencias del peso de 150 personas que usaron las canastillas para subir a esquiar cierto día. Construya un histograma con estos datos.

Clase	Frecuencia	Clase	Frecuencia
75- 89	10	150-164	23
90-104	11	165-179	9
105-119	23	180-194	9
120-134	26	195-209	6
135-149	31	210-224	2

- a) ¿Qué puede observar en este histograma acerca de los datos que no era aparente en la distribución de frecuencias?
- b) Si la capacidad de cada canastilla es dos personas y no más de 400 libras de peso total seguro, ¿qué puede hacer el operador para maximizar la capacidad de personas en las canastillas sin exceder la capacidad de peso seguro de una canastilla? ¿Los datos apoyan su propuesta?
- EA 2-6** El Central Carolina Hospital tiene los siguientes datos que representan el peso en libras de 200 bebés prematuros al momento de su nacimiento.

Clase	Frecuencia	Clase	Frecuencia
0.5-0.9	10	2.5-2.9	29
1.0-1.4	19	3.0-3.4	34
1.5-1.9	24	3.5-3.9	40
2.0-2.4	27	4.0-4.4	17

Construya una ojiva que le ayude a responder las siguientes preguntas:

- a) ¿Cuál es el valor medio aproximado en el conjunto de datos original?
- b) Si los bebés prematuros de menos de 3.0 libras se mantienen en una incubadora durante varios días como precaución, ¿cuál es el porcentaje aproximado de bebés prematuros en el Carolina que necesitarán una incubadora?

Aplicaciones

- **2-34** La siguiente distribución de frecuencias representa la duración de llamadas telefónicas efectuadas por 175 personas durante el fin de semana del Día del Trabajo. Construya un histograma para estos datos.

Duración en minutos	Frecuencia
1- 7	45
8-14	32
15-21	34
22-28	22
29-35	16
36-42	12
43-49	9
50-56	5

- a) Describa la forma general del histograma. ¿Puede observarse un patrón?
- b) Suponga que todas las personas hacían llamadas desde una habitación con 10 teléfonos y que cada persona sabía a qué clase de tiempo pertenecería la llamada. Sugiera un orden para que todas las llamadas terminen tan rápido como sea posible.
- c) ¿Afecta el orden a la duración de las llamadas terminadas?
- **2-35** Golden Acres es una asociación de propietarios de casas que opera un parque para casas móviles en las afueras de Orlando, Florida, donde los retirados tienen sus casas de invierno. Además de la renta de lotes, Golden Acres cobra una cuota mensual de \$12 para el uso de las instalaciones de actividades sociales de la casa club. Un miembro del consejo directivo ha notado que muchos de los residentes de mayor edad

nunca asisten a los eventos de la casa club, y ha propuesto exentar la cuota para los miembros de la asociación mayores de 60 años. Una investigación de 25 residentes dio los siguientes resultados de edades:

66	65	96	80	71
93	66	96	75	61
69	61	51	84	58
73	77	89	69	92
57	56	55	78	96

Construya una ojiva que le ayude a responder a las siguientes preguntas:

- En términos generales, ¿qué proporción de residentes sería elegible para la exención de la cuota?
 - ¿Qué cantidad aproximada tendría que cobrar la administración a los residentes que sí pagarían la cuota para cubrir el mismo costo total de operar la casa club?
- **2-36** Homero Willis, capitán de un barco pesquero de Salter Path, Carolina del Norte, tiene la creencia de que la pesca mínima para recuperar la inversión debe ser de 5,000 libras por viaje. A continuación tenemos los datos de una muestra de la pesca de 20 salidas al mar que el barco de Homero hizo recientemente:

6,500	6,700	3,400	3,600	2,000
7,000	5,600	4,500	8,000	5,000
4,600	8,100	6,500	9,000	4,200
4,800	7,000	7,500	6,000	5,400

Construya una ojiva que le ayude a responder las preguntas siguientes:

- Aproximadamente, ¿qué fracción de los viajes recupera exactamente la inversión según Homero?
 - ¿Cuál es el valor medio aproximado del arreglo de datos para los viajes del capitán?
 - ¿Qué pescas del barco de Homero exceden 80% del tiempo?
- **2-37** La organización Massachusetts Friends of Fish cuenta con los siguientes datos de contaminantes (en partes por millón) en 150 sitios del estado:

Contaminantes (en ppm)	Frecuencia	Contaminantes (en ppm)	Frecuencia
5.0- 8.9	14	25.0-28.9	16
9.0-12.9	16	29.0-32.9	9
13.0-16.9	28	33.0-36.9	7
17.0-20.9	36	37.0-40.9	4
21.0-24.9	20		

Construya una ojiva que le ayude a responder las siguientes preguntas:

- ¿Abajo de qué valor (aproximado) está la cuarta parte menor de estas observaciones?
 - Si los integrantes de la organización supervisan de cerca todos los sitios con más de 30 ppm de contaminantes, ¿qué porcentaje de sitios tendrá que supervisar?
- **2-38** Antes de construir una presa en el río Colorado, el Cuerpo de ingenieros del ejército de Estados Unidos realizó una serie de pruebas para medir el flujo de agua que pasa por el lugar de la presa. Los resultados de la prueba se utilizaron para construir la siguiente distribución de frecuencias:

Flujo de agua (miles de galones/min)	Frecuencia
1,001-1,050	7
1,051-1,100	21
1,101-1,150	32
1,151-1,200	49
1,201-1,250	58
1,251-1,300	41
1,301-1,350	27
1,351-1,400	11
Total	246

- Utilice los datos de la tabla para construir una distribución de frecuencias acumuladas “mayor que” y su ojiva.
- Utilice los datos de la tabla para construir una distribución de frecuencias acumuladas “menor que” y su ojiva.
- Utilice las ojivas que construyó para estimar qué proporción del flujo ocurre a menos de 1,300 millares de galones por minuto.

- **2-39** Pamela Mason, asesora de una pequeña empresa local de corretaje, intenta diseñar programas de inversión atractivos para jubilados. Ella sabe que si un inversionista potencial pudiera obtener cierto nivel de intereses, estaría dispuesto a invertir su capital, y que debajo de cierto nivel de intereses, no estaría dispuesto a hacerlo. De un grupo de 50 individuos, Pamela obtuvo los datos siguientes respecto a los diferentes niveles de intereses requeridos por cada individuo para invertir \$1,000:

Punto de indiferencia	Frecuencia	Punto de indiferencia	Frecuencia
\$70-74	2	\$90- 94	11
75-79	5	95- 99	3
80-84	10	100-104	3
85-89	14	105-109	2

- a) Construya distribuciones de frecuencias relativas acumuladas “menor que” y “mayor que”.
- b) Grafique las dos distribuciones del inciso a) como ojivas de frecuencias relativas.
- **2-40** En la redacción de un diario, se registró durante 50 días el tiempo requerido para formar la primera plana. Los datos, redondeados a la décima de minuto más cercana, se dan a continuación:

20.8	22.8	21.9	22.0	20.7	20.9	25.0	22.2	22.8	20.1
25.3	20.7	22.5	21.2	23.8	23.3	20.9	22.9	23.5	19.5
23.7	20.3	23.6	19.0	25.1	25.0	19.5	24.1	24.2	21.8
21.3	21.5	23.1	19.9	24.2	24.1	19.8	23.9	22.8	23.9
19.7	24.2	23.8	20.7	23.8	24.3	21.1	20.9	21.6	22.7

- a) Ordene los datos en un arreglo ascendente.
- b) Construya una distribución de frecuencias absolutas y una distribución de frecuencias acumuladas “menor que” a partir de los datos. Utilice intervalos de 0.8 minutos.
- c) Construya un polígono de frecuencias con base en los datos.
- d) A partir de los datos, construya una ojiva “menor que”.
- e) Tomando en cuenta su ojiva, estime qué porcentaje del tiempo puede formarse la primera plana en menos de 24 minutos.
- **2-41** Chien-Ling Lee, posee una tienda de discos especializada en grabaciones de voz. Lee tiene 35 meses de datos de ventas brutas, arreglados en una distribución de frecuencias:

Venta mensual	Frecuencia	Venta mensual	Frecuencia
\$10,000-12,499	2	\$20,000-22,499	6
12,500-14,999	4	22,500-24,999	8
15,000-17,499	7	25,000-27,499	2
17,500-19,999	5	27,500-29,999	1

- a) Construya una distribución de frecuencias relativas.
- b) Construya, sobre una misma gráfica, un histograma de frecuencias relativas y un polígono de frecuencias relativas.
- **2-42** La Asociación nacional de vendedores de bienes raíces de Estados Unidos recabó los datos siguientes de una muestra de 130 vendedores, que representan sus ingresos totales por comisiones anuales:

Ingresos	Frecuencia
\$ 5,000 o menos	5
5,001-10,000	9
10,001-15,000	11
15,001-20,000	33
20,001-30,000	37
30,001-40,000	19
40,001-50,000	9
Más de 50,000	7

Construya una ojiva que le ayude a responder las preguntas siguientes:

- a) ¿Aproximadamente qué proporción de vendedores gana más de \$25,000?
- b) ¿Alrededor de cuánto gana el vendedor “medio” de la muestra?
- c) ¿Cuál es el ingreso aproximado al año de un vendedor cuyo desempeño es cercano al 25% del ingreso máximo anual esperado?

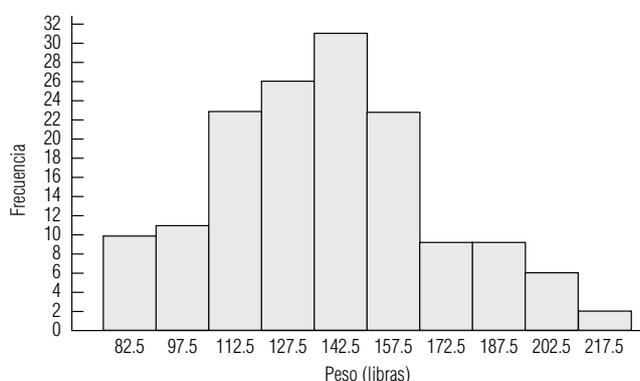
- 2-43** Springfield es un pueblo universitario con los problemas usuales de estacionamiento. El pueblo permite a las personas que tienen multas por estacionar sus autos en lugares prohibidos presentar su caso ante un oficial administrativo para ver si puede anularles la multa. El oficial administrativo de la ciudad recolectó la siguiente distribución de frecuencias para el tiempo dedicado a cada apelación:

Minutos dedicados a la apelación	Frecuencia	Minutos dedicados a la apelación	Frecuencia
Menos de 2	30	8- 9	70
2-3	40	10-11	50
4-5	40	12-13	50
6-7	90	14-15	30
			<u>400</u>

- Construya una distribución de frecuencias acumuladas “menor que”.
- Construya una ojiva basada en el inciso a).
- El administrador de la ciudad considerará simplificar el proceso de apelación si más de 50% de las apelaciones lleva más de 4 minutos. ¿Qué porcentaje toma más de 4 minutos? ¿Cuál es el tiempo aproximado para la apelación número 200 (punto medio)?

Soluciones a los ejercicios de autoevaluación

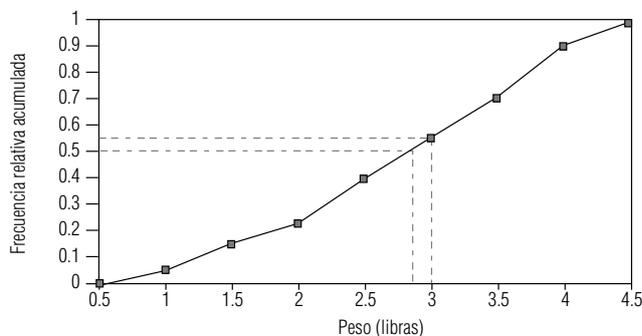
EA 2-5



- La cola inferior de la distribución es más ancha (tiene más observaciones) que la cola superior.
- Debido a que existen tan pocas personas que pesan 180 libras o más, el operador puede formar parejas de cada persona que se ve corpulenta con una persona delgada. Esto se puede hacer prácticamente sin retrasar los turnos individuales para subir a las canastillas.

EA 2-6

Clase	Frecuencia relativa acumulada	Clase	Frecuencia relativa acumulada
0.5-0.9	0.050	2.5-2.9	0.545
1.0-1.4	0.145	3.0-3.4	0.715
1.5-1.9	0.265	3.5-3.9	0.915
2.0-2.4	0.400	4.0-4.4	1.000



- El valor medio es alrededor de 2.8 libras.
- Cerca de 55% necesitará incubadora.

Loveland Computers

Caso 2: Arreglo de datos El año nuevo de 1995 sorprendió a Lee Azko observando una ligera capa de nieve en los suburbios de Denver a través de la ventana. Lee se había graduado antes del tiempo normal en la Universidad de Colorado, un semestre menos que el periodo acostumbrado de cuatro años, gracias a un puñado de créditos académicos adelantados desde el bachillerato. Lee se encontraba preocupado y emocionado al mismo tiempo, pues al día siguiente comenzaría a buscar, seriamente, un puesto de trabajo para un recién graduado con buena capacitación y poca experiencia en el mundo real.

La contemplación del futuro fue interrumpida por una llamada telefónica de su tío. “Iba a llamarte de todas formas para felicitarte por haber terminado la escuela un semestre antes. Pero hay otra razón más para hablarte: han surgido algunas cosas en la compañía y parece que también surge la necesidad urgente de alguien capaz de procesar algunos números. ¿Por qué no vienes a la oficina, mañana en la mañana, y te digo qué es lo que tengo en mente?”

Lee sabía que la compañía de su tío Walter, Loveland Computers, estaba creciendo a pasos agigantados. Walter Azko había desarrollado la compañía de computadoras a partir de una extraña base. A diferencia de Lee, Walter nunca terminó la universidad. “Estaba ganando mucho dinero como para quedarme en la escuela”, solía explicar. Walter había recorrido extensamente el Lejano Oriente con sus padres, de manera que parecía natural que iniciara un importante negocio de importaciones mientras aún era estudiante en Boulder. Importaba todo tipo de mercancía que pudiera venderse barata y que fuera atractiva para los estudiantes: muebles, regalos, utensilios para el hogar y algo de ropa.

A principios de la década de 1980, en uno de sus viajes de compras, le ofrecieron algunas computadoras personales. Mirando retrospectivamente, ahora le parecían horribles. No tenían mucha memoria y carecían de disco duro, pero eran en extremo baratas, y las vendió rápidamente a algunos fanáticos de la computación de la universidad. El negocio de las computadoras creció, y en un lapso de dos años, Walter vendió su empresa de importaciones al menudeo y se concentró exclusivamente en la importación y venta de computadoras.

Su primer paso consistió en rentar un edificio comercial en Loveland, Colorado, donde los alquileres eran mucho más baratos que en Boulder. Desde ese lugar, podía vender directamente a los estudiantes de las universidades de Boulder, Fort Collins y Greeley. Situado a casi (o menos) una hora del aeropuerto internacional Stapleton de Denver, Loveland era el lugar indicado para las importaciones que arribaran por vía aérea y un buen sitio para contratar trabajadores de medio tiempo. El nombre de Loveland Computers surgió como algo natural.

Al principio, Walter Azko actuó como su propio agente de ventas: entregaba personalmente las máquinas, transportándolas en la parte trasera de su automóvil. Walter logró sus ventas sobre todo gracias a un buen precio, y completó la publicidad de boca en boca con algunos anuncios en los periódicos universitarios. Parecía que él era el único en los alrededores que vendía directamente a los estudiantes y a los aficionados a la computación. El mercado de Walter parecía ser algo completamente distinto al mercado en el que habían puesto su atención los grandes gigantes de la industria. En la cima, IBM utilizaba una distribución al menudeo bastante costosa dirigida al segmento corporativo. Apple defendía su estrategia de precios altos con una paquetería fácil de operar con un “ratón”, que no era igualada por las computadoras personales compatibles con IBM.

Azko empezó a leer revistas sobre computación y se dio cuenta de que él no era el único que tenía una “tienda de cajas” (como se refiere la industria a las compañías que venden, a usuarios finales, computadoras empacadas en cajas sin ofrecer servicios adicionales o con muy pocos de éstos). Una o dos compañías habían encontrado proveedores baratos en el extranjero y estaban desarrollando una estrategia de ventas por correo. Walter pensó que los clientes no iban a sentirse animados a comprar equipo tan caro y novedoso, sin conocerlo, pero la llegada de un nuevo embarque de computadoras con disco duro preinstalado le dio la motivación para colocar unos cuantos anuncios de su compañía.

Así pues, Loveland Computers alcanzó la categoría de las tiendas de cajas con pedidos por correo a nivel nacional, y para 1988, la compañía era una de las dos docenas que constituían este mercado. En conjunto, las compañías de pedidos por correo poseían aproximadamente el mismo porcentaje de mercado que la “Big Blue” (IBM): aproximadamente 20%. Pero el mercado para las PC era enorme y estaba creciendo con rapidez. En 1990, Loveland Computers registraba ventas de \$10 millones al trimestre, aún a precios de descuento, con ganancias que regularmente representaban el 6% de las ventas. El tío Walter se había convertido en un hombre rico.

Con la experiencia, Walter Azko se dio cuenta de que para dar al cliente exactamente lo que deseaba, existían ventajas si armaba las máquinas en sus instalaciones, cada vez más grandes. Jamás se consideró un fabricante; sólo un ensamblador de partes prefabricadas, como controladores de unidades de disco y fuentes de poder, pero gracias a sus contactos con fabricantes del extranjero, Walter era capaz de buscar siempre los mejores precios para poder mantenerlos bajos.

Para configurar máquinas nuevas y ayudar con las especificaciones, Walter contrató a una joven y brillante ingeniera, Gracia Delaguardia. Ella conocía el hardware, pues había llevado a la práctica varios proyectos de desarrollo para la empresa Storage Technology. En unos cuantos años en Loveland Computers, Delaguardia formó un equipo de desarrollo de más de dos docenas de personas y su recompensa fue convertirse en socia de la empresa.

Loveland Computers tuvo unos cuantos reveses debido a una estimación errónea de la demanda. Walter Azko se sentía siempre optimista con respecto a las ventas, así que el inventario de los componentes siempre era mucho mayor de lo necesario. En una o dos ocasiones ocurrieron penosos desperdicios, como cuando todo un embarque de fuentes de poder resultó inútil, ya que generaban una corriente demasiado baja para el último modelo Loveland. Gracia Delaguardia llegó a la conclusión de que Loveland debía ser capaz de administrar mejor sus suministros, pero parecía difícil predecir el comportamiento del mercado con un mes de anticipación.

Después de pasar una noche de insomnio, Lee Azko se encontró con el fundador y presidente de Loveland Computers. “Ven, siéntate acá junto a la ventana, puedes ver mi nuevo automóvil deportivo, un Mercedes 500 SL”, le dijo Walter Azko al darle la bienvenida a su joven visitante. Déjame plantearte el problema. Tú sabes que las cosas se mueven demasiado aprisa en este lugar. Parece que un modelo dura alrededor de seis meses y luego tenemos que reemplazarlo con algo más complejo. Hasta este punto, he dependido, más o menos, del banco local en cuanto a financiamiento. Pero es un buen negocio y estamos llamando la atención de algunas personas de Wall Street. Tal vez logremos una “colocación privada”, ésa es la forma en que podemos obtener capital de uno o dos inversionistas o banqueros adinerados para expandirnos, y luego, más adelante, quizá sea factible cotizar ac-

ciones de la compañía en la bolsa. La cuestión es que desean saber más cosas sobre nuestro crecimiento de ventas: qué tanto proviene de cual o tal producto, cosas como ésas. Desean saber cuánto tiempo dura cada modelo, qué es lo que debemos proyectar para el año siguiente. En este momento, por supuesto, tengo informes mensuales sobre las ventas que se remontan casi al inicio de la empresa. La buena noticia es que todo está en discos; la mala es que nos pasamos cambiando formatos, de modo que resulta demasiado difícil comparar cifras. Y, por supuesto, nadie tiene ganas de sumergirse en, digamos, 48 meses de informes. Tu trabajo consiste en organizar toda la información, de modo que tenga sentido para cuando los estafadores de la ciudad vengan al pueblo en su jet privado.

“¿Cuándo debo empezar, tío?”, preguntó Lee Azko, que se vio tomado por sorpresa ante la tarea que tenía enfrente.

“Ya empezaste, le respondió Walter. Lo importante es cuándo vas a terminar. Esas personas estarán aquí el próximo lunes.”

Lee se despidió de un paseo a esquiar durante el fin de semana, sacó una libreta y empezó a bosquejar un plan de acción.

Preguntas de estudio: ¿Qué información deberá reunir, además de la financiera relativa a ventas e ingresos? ¿Qué formato reflejará con mayor claridad el rápido crecimiento de la compañía en una presentación de negocios de 45 minutos?



Ejercicio de base de datos computacional

HH Industries

“Atención todos, ella es Laurel. Laurel McRae”, anunció Hal Rodgers, presidente de HH Industries, en la reunión semanal de personal. “Laurel, ellos son Stan Hutchings, vicepresidente de ventas; Peggy Noble, gerente de contabilidad y procesamiento de datos; Bob Ritchie, gerente de compras y control de inventario, y Gary Russell, gerente de operaciones.

“Todos ustedes saben que HH Industries va por buen camino”, continuó Hal. “Los últimos tres años hemos tenido estabilidad y un prometedor crecimiento en cierto número de áreas. Sin embargo, por lo regular, basamos nuestras decisiones en los años que tenemos de experiencia colectiva y en las buenas corazonadas. Laurel es una experimentada analista de datos y planeadora estratégica, y se ha integrado a nuestro equipo para ayudarnos a analizar, de manera más cuantitativa y estadística, dónde estamos ahora y dónde esperamos estar dentro de algunos años. Podemos ser buenos, pero las complejas investigaciones de mercado y la estrategia de análisis muestran que tendremos un gran futuro. Además, tal vez po-

damos, por fin, ¡hallar un uso productivo para algunos de los documentos que generamos!”

El personal rió entre dientes. Si la compañía no prosperaba, no era debido a la falta de datos. Desde la introducción de un programa de procesamiento de datos que se hizo especialmente para la empresa, en el año fiscal anterior, se tenía disponible una gran cantidad de datos, algunos útiles y otros bastante confusos. Las cifras sobre ventas diarias y márgenes de utilidad se almacenaban religiosamente, junto con los datos detallados acerca de inventarios y embarques. Nadie tenía idea aún de qué se podría hacer con ellos, aunque el presidente y su personal directivo mantenían los registros como simples cifras importantes.

De vuelta en su oficina, Laurel se puso a analizar su reciente cambio a las oficinas generales de HH Industries en Florida Suncoast, desde su antiguo puesto en la fábrica de juguetes Cold River en las Montañas Rocallosas. No estaba muy segura sobre el uso que daba el presidente de HH a la palabra “experimentada”, pero ella había hecho lo mejor que había podido. La decisión de abandonar al exitoso fabricante de trineos y juguetes había sido difícil, pero tenía confianza en que las compañías de almacenamiento y distribución, como HH Industries, eran una sólida apuesta para el futuro. Y, durante la entrevista inicial, Laurel quedó impresionada con

Hal Rodgers y con el ambiente positivo y eficiente de la empresa. Pronto se daría cuenta de si le iba a gustar la industria hidráulica tanto como la de juguetes.

“Conócenos”, le había pedido Hal. “El personal directivo está completamente a tu disposición. Haz preguntas; échale un vistazo a los datos que tenemos. No sé exactamente en qué pueda ayudarnos la estadística, pero tengo plena confianza en ti. Llegaste bastante bien recomendada como analista y como pensadora con iniciativa.”

“Bueno —pensó Laurel—, aquí vamos.” Primera parada, una tarde con Stan Hutchings para ponerse un poco al tanto de lo que sucedía en la compañía. Ella sabía que Stan tenía trabajando en HH Industries más tiempo que cualquier otro miembro del personal directivo y que poseía una excelente intuición para la industria hidráulica.

Días más tarde y después de varias de esas juntas de familiarización, los datos empezaron a amontonarse en su escritorio vacío. Laurel reflexionó sobre lo que había aprendido acerca de la empresa. HH Industries era un típico negocio familiar, establecido hacía más de 20 años por la familia Douglas. Handy Hydraulics (como se le conocía en aquel entonces) surgió para cubrir una necesidad percibida por sus fundadores: una fuente de refacciones y reparación para la creciente industria hidráulica portátil. La próspera población de la década de 1960 requería el apoyo de un número cada vez mayor de vehículos para la construcción, camiones de basura y otras piezas grandes de equipo, que a su vez requerían refacciones y reparación para una enorme variedad de sellos hidráulicos, bombas, cilindros, medidores y demás. En su papel de distribuidor, Handy Hydraulics localizaba proveedores de partes y vendía directamente, con el nombre del fabricante, o empacaba partes sueltas en paquetes de reparación y las revendía con su propio nombre.

Durante los primeros cinco años del negocio hubo un crecimiento estable, aunque se hizo poca investigación de mercado. La publicidad de boca en boca y un importante mercado constituyeron un saludable entorno para la compañía. Las primeras ventas fueron casi todas en Florida y no fue sino hasta después de producir el primer catálogo, en 1974, que el negocio empezó a expandirse hacia el norte, a Alabama y Georgia.

Una comercialización de “fuerza bruta” era el siguiente paso y Laurel sonrió al pensar en la pobre secretaria que tenía que enviar correspondencia a los futuros clientes, seleccionados en las páginas de la sección amarilla de todas las comunidades, en todo el país; es decir, más de 25,000 personas. La filosofía era sencilla: en los lugares donde hay grandes concentraciones de población, hay también camiones para basura y equipo de construcción que prestan apoyo a las comunidades. Y funcionó. A finales de la década de los setenta y principio de los ochenta, la compañía tuvo un crecimiento sostenido, cada día tenían nuevos clientes. Desafortunadamente, y algo que resulta típico en las compañías familiares, la administración no pudo mantener el paso.

Para ese entonces, habían surgido numerosos competidores a todo lo largo y ancho de Estados Unidos, algunos de los cuales habían sido, en un principio, clientes de Handy Hydraulics. Se hizo evidente que el propósito de la compañía de mantener su presencia a nivel nacional sólo podría cumplirse abriendo almacenes satélite en algún otro lugar, para poder seguir prestando el servicio de entrega al día siguiente, a un costo razonable, a todas las áreas del país. Con este propósito, la familia Douglas vendió Handy Hydraulics a la compañía actual, BMP Enterprises, y el señor Douglas firmó un contrato de tres años para seguir fungiendo como presidente. Con el capital adicional que aportó la compañía inversionista se abrieron almacenes en Arizona (1985) y Ohio (1986). Sin embargo, la compañía fue mantenida sin mucho movimiento por su fundador original, quien prestó poca atención a la forma de administrar mejor los almacenes satélite. De manera similar, no se reconoció la importancia del cambiante medio comercial (competencia creciente, nuevas tecnologías y estrategias de administración disponibles). El resultado fue un negocio fuera de control, que se sofocaba a sí mismo con políticas y procedimientos que habían funcionado y que ahora resultaban demasiado rígidos. Algo tenía que suceder.

Y sucedió cuando el señor Douglas se retiró en 1988 y BMP Enterprises llevó a Hal Rodgers para que intentara salvar a Handy Hydraulics. Se trataba de un sólido ejecutivo de negocios con buena intuición y un todavía mejor “don de gentes”. Hal heredó una compañía en serios problemas. Aunque tenía ingresos de \$900,000 por ventas trimestrales, una nómina y gastos de operación en extremo altos hacían que hubiera pérdidas netas.

Durante los tres años siguientes, se introdujeron cambios significativos que lograron aumentar las ventas mientras se disminuían los costos. La nómina fue recortada al mínimo e incluso se cerró un local de compra directa, que alguna vez fue útil para las relaciones públicas pero que después se convirtió en una carga costosa. Se instalaron números telefónicos para recibir pedidos sin costo para el cliente. El almacén de Ohio fue cerrado y, casi dos años más tarde, se abrió una versión moderna de éste en Pennsylvania. El catálogo de la compañía, que antes consistía de un voluminoso montón de hojas unidas por un soporte de tres aros y que tenía que actualizarse mediante continuos envíos por correo, fue reducido a una versión de amplia distribución que representaba de manera más clara y concisa los productos de la compañía. Por último, para publicitar y celebrar la nueva imagen de la compañía, se le cambió el nombre a HH Industries.

Ésta fue la organización que encontró Laurel a su llegada. Hizo un resumen de la estructura actual: tres centros de distribución (Florida, Arizona y Pennsylvania); tres líneas de productos (sellos y equipos para sellos, equipo pesado terminado —cilindros, bombas, válvulas, etc.— y refacciones y servicio de reparación). La compañía tenía 42 empleados de tiempo completo y nueve de medio tiempo, más de 3,000

cuentas activas de clientes y aproximadamente 15,000 productos de línea en existencia. El año fiscal corría de diciembre a noviembre y en cada trimestre, ahora, las cifras de ventas se acercaban en promedio a \$1.4 millones. “¡Guau!” pensó Laurel. “Un tanto diferente de la fábrica de juguetes. Pero me van a pagar para que haga estadística y análisis, de manera que veamos si puedo clavarle el diente a este monstruo.”

Laurel extrajo los datos de ventas anuales más actualizados (correspondientes al tercer y cuarto trimestres de 1990, y al primero y segundo de 1991); tanto el número de pedidos por día como el valor en dólares de tales pedidos (referidos como “ventas”), por centro de distribución. Los archivos CH02.xxx del disco de datos contienen esta información. Por lo que Laurel pudo observar, el ambiente completo de la compañía parecía girar alrededor de lo que llamaron “cifras del día”: ventas conjuntas totales por día. Sin embargo, la experiencia de Lau-

rel le dijo que tenía que buscar con más profundidad. Sabía, por ejemplo, que las ventas diarias en dólares eran producto directo de dos factores: el número de pedidos por día y el valor promedio en dólares por cada pedido.

1. Construya histogramas y distribuciones de frecuencias relativas para el *tamaño promedio de pedidos por día* (ventas totales divididas entre el total de pedidos) para los últimos cuatro trimestres. En cada uno utilice anchos de intervalo de 20 y tome el primer intervalo de cero a 20.
2. Construya diagramas parecidos, por trimestre, para el *número total de pedidos por día*. Utilice anchos de intervalo de 10 con el primer intervalo de 100 a 110.
3. ¿Qué patrones de cambio se hacen notorios en los datos de un trimestre a otro? ¿Cuál sería la explicación posible de tal comportamiento?

Repaso del capítulo

● Términos introducidos en el capítulo 2

Arreglo u ordenamiento de datos Organización de los datos sin procesar en orden descendente o ascendente.

Clase de extremo abierto Clase que permite que el extremo superior o inferior de un esquema de clasificación cuantitativo no tenga límite.

Conjunto de datos Una colección de datos.

Curva de frecuencias Polígono de frecuencias suavizado al aumentar el número de clases y datos puntuales a un conjunto de datos.

Dato puntual Una sola observación de un conjunto de datos.

Datos Colección de cualquier número de observaciones relacionadas de una o más variables.

Datos continuos Datos que pueden pasar de una clase a la siguiente sin interrupción y que pueden expresarse mediante números enteros o fraccionarios.

Datos discretos Datos que no pasan de una clase a la siguiente sin que haya una interrupción; esto es, en donde las clases representan categorías o cuentas distintas que pueden representarse mediante números enteros.

Datos sin procesar Los datos antes de ser organizados o analizados por métodos estadísticos.

Distribución de frecuencias Presentación organizada de datos que muestra el número de observaciones del conjunto de datos que entra en cada una de las clases de un conjunto de clases mutuamente excluyentes y colectivamente exhaustivas.

Distribución de frecuencias acumuladas Presentación de datos en forma de tabla que muestra cuántos datos están por encima o por debajo de ciertos valores.

Distribución de frecuencias relativas Presentación de un conjunto de datos en el que se muestra la fracción o porcentaje del total del conjunto de datos que entra en cada clase mutuamente excluyente y colectivamente exhaustiva.

Histograma Gráfica de un conjunto de datos compuesta por una serie de rectángulos, cada uno con un ancho proporcional al rango de los valores de cada clase y altura proporcional al número de elementos que entran en la clase (o altura proporcional a la fracción de elementos de la clase).

Muestra Colección de algunos elementos —no todos—, de la población bajo estudio, utilizada para describir poblaciones.

Muestra representativa Muestra que contiene las características importantes de la población en las mismas proporciones en que están contenidas en la población.

Ojiva Gráfica de una distribución de frecuencias acumuladas.

Población Colección de todos los elementos que se están estudiando y sobre los cuales intentamos llegar a conclusiones.

Polígono de frecuencias Línea que une los puntos medios de cada clase de un conjunto de datos, trazada a la altura correspondiente a la frecuencia de los datos.

Diagrama de tallo y hojas

Los datos estadísticos obtenidos de poblaciones grandes pueden ser muy útiles para estudiar el comportamiento de la distribución si se presentan en una combinación tabular y gráfica conocida como **diagrama de tallo y hojas**.

Para ejemplificar la elaboración de un diagrama de tallo y hojas considere los datos de la tabla 1.4, que especifican la “vida” de 40 baterías para automóvil similares, registradas al décimo de año más cercano. Las baterías se garantizan por tres años. Comience por dividir cada observación en dos partes: una para el tallo y otra para las hojas, de manera que el tallo represente el dígito entero que antecede al decimal y la hoja corresponda a la parte decimal del número. En otras palabras, para el número 3.7 el dígito 3 se designa al tallo y el 7 a la hoja. Para nuestros datos los cuatro tallos 1, 2, 3 y 4 se listan verticalmente del lado izquierdo de la tabla 1.5, en tanto que las hojas se registran en el lado derecho correspondiente al valor del tallo adecuado. Entonces, la hoja 6 del número 1.6 se registra enfrente del tallo 1; la hoja 5 del número 2.5 enfrente del tallo 2; y así sucesivamente. El número de hojas registrado junto a cada tallo se anota debajo de la columna de frecuencia.

Tabla 1.4: Vida de las baterías para automóvil

2.2	4.1	3.5	4.5	3.2	3.7	3.0	2.6
3.4	1.6	3.1	3.3	3.8	3.1	4.7	3.7
2.5	4.3	3.4	3.6	2.9	3.3	3.9	3.1
3.3	3.1	3.7	4.4	3.2	4.1	1.9	3.4
4.7	3.8	3.2	2.6	3.9	3.0	4.2	3.5

Tabla 1.5: Diagrama de tallo y hojas de la vida de las baterías

Tallo	Hoja	Frecuencia
1	69	2
2	25669	5
3	0011112223334445567778899	25
4	11234577	8

El diagrama de tallo y hojas de la tabla 1.5 contiene sólo cuatro tallos y, en consecuencia, no ofrece una representación adecuada de la distribución. Para solucionar este problema es necesario aumentar el número de tallos en nuestro diagrama. Una manera sencilla de hacerlo consiste en escribir dos veces cada valor del tallo y después registrar las hojas 0, 1, 2, 3 y 4 enfrente del valor del tallo adecuado, donde aparezca por primera vez; y las hojas 5, 6, 7, 8 y 9 enfrente de este mismo valor del tallo, donde aparece la segunda vez. El diagrama doble de tallo y hojas modificado se ilustra en la tabla 1.6, donde los tallos que corresponden a las hojas 0 a 4 fueron codificados con el símbolo ★ y los tallos correspondientes a las hojas 5 a 9 con el símbolo •.

En cualquier problema dado debemos decidir cuáles son los valores del tallo adecuados. Esta decisión se toma hasta cierto punto de manera arbitraria, aunque debemos guiarnos por el tamaño de nuestra muestra. Por lo general elegimos entre 5 y 20 tallos. Cuanto más pequeña sea la cantidad de datos disponibles, más pequeña será nuestra elección del número de tallos. Por ejemplo, si los datos constan de números del 1 al 21,

los cuales representan el número de personas en la fila de una cafetería en 40 días laborales seleccionados al azar, y elegimos un diagrama doble de tallo y hojas, los tallos serían 0★, 0•, 1★, 1• y 2★, de manera que la observación de 1 más pequeña tiene tallo 0★ y hoja 1, el número 18 tiene tallo 1• y hoja 8, y la observación de 21 más grande tiene tallo 2★ y hoja 1. Por otro lado, si los datos constan de números de \$18,800 a \$19,600, que representan las mejores ventas posibles de 100 automóviles nuevos, obtenidos de cierto concesionario, y elegimos un diagrama sencillo de tallo y hojas, los tallos serían 188, 189, 190, ..., 196 y las hojas contendrían ahora dos dígitos cada una. Un automóvil que se vende en \$19,385 tendría un valor de tallo de 193 y 85 en los dos dígitos de la hoja. En el diagrama de tallo y hojas, las hojas de dígitos múltiples que pertenecen al mismo tallo por lo regular están separadas por comas. En los datos generalmente se ignoran los puntos decimales cuando todos los números a la derecha del punto decimal representan hojas, como en el caso de las tablas 1.5 y 1.6. Sin embargo, si los datos constaran de números que van de 21.8 a 74.9, podríamos elegir los dígitos 2, 3, 4, 5, 6 y 7 como los tallos, de manera que un número como 48.3 tendría un valor de tallo de 4 y un valor de hoja de 8.3.

Tabla 1.6: Diagrama doble de tallo y hojas para la vida de las baterías

Tallo	Hoja	Frecuencia
1•	69	2
2★	2	1
2•	5669	4
3★	001111222333444	15
3•	5567778899	10
4★	11234	5
4•	577	3

El diagrama de tallo y hojas representa una manera eficaz de resumir los datos. Otra forma consiste en el uso de la **distribución de frecuencias**, donde los datos, agrupados en diferentes clases o intervalos, se pueden construir contando las hojas que pertenecen a cada tallo y considerando que cada tallo define un intervalo de clase. En la tabla 1.5 el tallo 1 con 2 hojas define el intervalo 1.0-1.9, que contiene 2 observaciones; el tallo 2 con 5 hojas define el intervalo 2.0-2.9, que contiene 5 observaciones; el tallo 3 con 25 hojas define el intervalo 3.0-3.9, con 25 observaciones; y el tallo 4 con 8 hojas define el intervalo 4.0-4.9, que contiene 8 observaciones. Para el diagrama doble de tallo y hojas de la tabla 1.6 los tallos definen los siete intervalos de clase 1.5-1.9, 2.0-2.4, 2.5-2.9, 3.0-3.4, 3.5-3.9, 4.0-4.4 y 4.5-4.9, con frecuencias 2, 1, 4, 15, 10, 5 y 3, respectivamente.

Histograma

Al dividir cada frecuencia de clase entre el número total de observaciones, obtenemos la proporción del conjunto de observaciones en cada una de las clases. Una tabla que lista las frecuencias relativas se denomina **distribución de frecuencias relativas**. En la tabla 1.7 se presenta la distribución de frecuencias relativas para los datos de la tabla 1.4, que muestra los puntos medios de cada intervalo de clase.

La información que brinda una distribución de frecuencias relativas en forma tabular es más fácil de entender si se presenta en forma gráfica. Con los puntos medios de

Tabla 1.7: Distribución de frecuencias relativas de la vida de las baterías

Intervalo de clase	Punto medio de la clase	Frecuencia, f	Frecuencia relativa
1.5–1.9	1.7	2	0.050
2.0–2.4	2.2	1	0.025
2.5–2.9	2.7	4	0.100
3.0–3.4	3.2	15	0.375
3.5–3.9	3.7	10	0.250
4.0–4.4	4.2	5	0.125
4.5–4.9	4.7	3	0.075

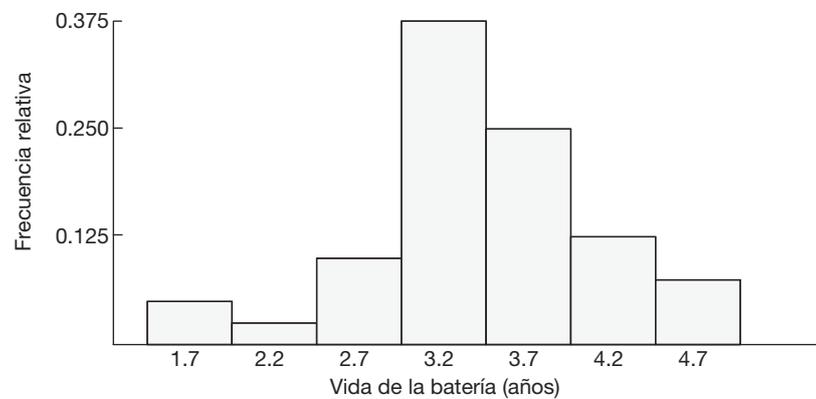


Figura 1.6: Histograma de frecuencias relativas.

cada intervalo y las frecuencias relativas correspondientes construimos un **histograma de frecuencias relativas** (figura 1.6).

Muchas distribuciones de frecuencias continuas se pueden representar gráficamente mediante la curva en forma de campana característica de la figura 1.7. Herramientas gráficas como las de las figuras 1.6 y 1.7 ayudan a comprender la naturaleza de la población. En los capítulos 5 y 6 examinaremos una propiedad de la población que se conoce como **distribución**. Aunque más adelante en este texto se proporcionará una definición más precisa de una distribución o de una **distribución de probabilidad**, aquí podemos visualizarla como la que se podría haber visto en el límite de la figura 1.7 cuando el tamaño de la muestra aumentara.

Se dice que una distribución es **simétrica** si se puede doblar a lo largo de un eje vertical de manera que ambos lados coincidan. Si una distribución carece de simetría respecto de un eje vertical, se dice que está **sesgada**. La distribución que se ilustra en la figura 1.8a se dice que está sesgada a la derecha porque tiene una cola derecha larga y una cola izquierda mucho más corta. En la figura 1.8b observamos que la distribución es simétrica; mientras que en la figura 1.8c está sesgada a la izquierda.

Al girar un diagrama de tallo y hojas en dirección contraria a la de las manecillas del reloj en un ángulo de 90° , vemos que las columnas de hojas que resultan forman una imagen parecida a un histograma. Por lo tanto, si nuestro objetivo principal al observar los datos es determinar la forma general o la forma de la distribución, rara vez será necesario construir un histograma de frecuencias relativas.

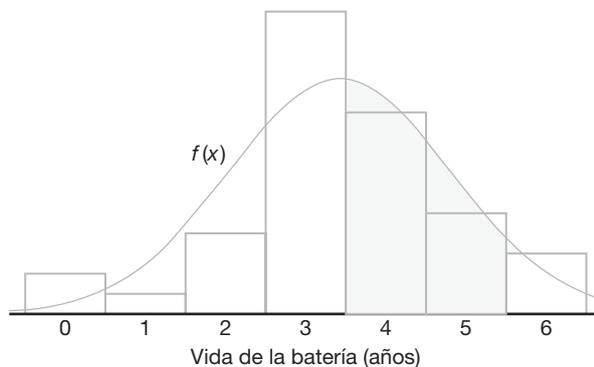


Figura 1.7: Estimación de la distribución de frecuencias.

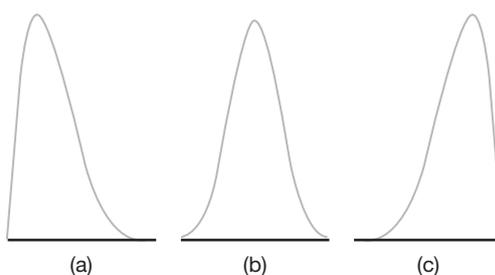


Figura 1.8: Sesgo de los datos.

Gráfica de caja y bigote o gráfica de caja

Otra presentación que es útil para reflejar propiedades de una muestra es la **gráfica de caja y bigote**, la cual encierra el *rango intercuartil* de los datos en una caja que contiene la mediana representada. El rango intercuartil tiene como extremos el percentil 75 (cuartil superior) y el percentil 25 (cuartil inferior). Además de la caja se prolongan “bigotes”, que indican las observaciones alejadas en la muestra. Para muestras razonablemente grandes la presentación indica el centro de localización, la variabilidad y el grado de asimetría.

Además, una variación denominada **gráfica de caja** puede ofrecer al observador información respecto de cuáles observaciones son **valores extremos**. Los valores extremos son observaciones que se consideran inusualmente alejadas de la masa de datos. Existen muchas pruebas estadísticas diseñadas para detectar este tipo de valores. Técnicamente se puede considerar que un valor extremo es una observación que representa un “evento raro” (existe una probabilidad pequeña de obtener un valor que esté lejos de la masa de datos). El concepto de valores extremos volverá a surgir en el capítulo 12 en el contexto del análisis de regresión.

La información visual en las gráficas de caja y bigote o en las de caja no intenta ser una prueba formal de valores extremos, más bien se considera una herramienta de diagnóstico. Aunque la determinación de cuáles observaciones son valores extremos varía de acuerdo con el tipo de software que se emplee, un procedimiento común para determinarlo consiste en utilizar un **múltiplo del rango intercuartil**. Por ejemplo, si la distancia desde la caja excede 1.5 veces el rango intercuartil (en cualquier dirección), la observación se podría considerar un valor extremo.

Ejemplo 1.5: Se midió el contenido de nicotina en una muestra aleatoria de 40 cigarrillos. Los datos se presentan en la tabla 1.8.

Tabla 1.8: Valores de nicotina para el ejemplo 1.5

1.09	1.92	2.31	1.79	2.28	1.74	1.47	1.97
0.85	1.24	1.58	2.03	1.70	2.17	2.55	2.11
1.86	1.90	1.68	1.51	1.64	0.72	1.69	1.85
1.82	1.79	2.46	1.88	2.08	1.67	1.37	1.93
1.40	1.64	2.09	1.75	1.63	2.37	1.75	1.69

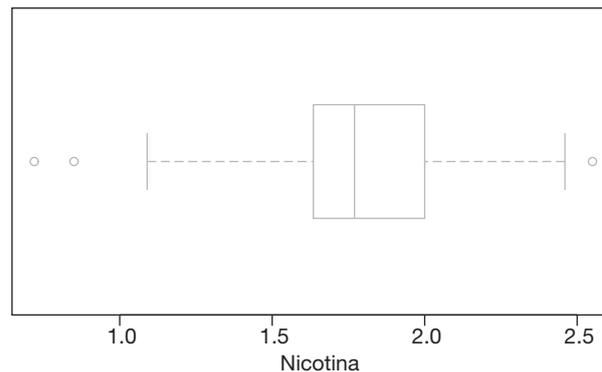


Figura 1.9: Gráfica de caja y bigote para el ejemplo 1.5.

La figura 1.9 muestra la gráfica de caja y bigote de los datos, la cual describe las observaciones 0.72 y 0.85 como valores extremos moderados en la cola inferior; en tanto que la observación 2.55 es un valor extremo moderado en la cola superior. En este ejemplo el rango intercuartil es 0.365, y 1.5 veces el rango intercuartil es 0.5475. Por otro lado, la figura 1.10 presenta un diagrama de tallo y hojas. ─

Ejemplo 1.6: Considere los datos de la tabla 1.9, que constan de 30 muestras que miden el grosor de las “asas” de latas de pintura (véase el trabajo de Hogg y Ledolter de 1992 en la bibliografía). La figura 1.11 describe una gráfica de caja y bigote para este conjunto asimétrico de datos. Observe que el bloque izquierdo es considerablemente más grande que el bloque de la derecha. La mediana es 35. El cuartil inferior es 31, mientras que el superior es 36. Advierta también que la observación alejada de la derecha está más lejos de la caja que la observación extrema de la izquierda. No hay valores extremos en este conjunto de datos. ─

El punto decimal se encuentra 1 dígito(s) a la izquierda de I

```

7 | 2
8 | 5
9 |
10 | 9
11 |
12 | 4
13 | 7
14 | 07
15 | 18
16 | 3447899
17 | 045599
18 | 2568
19 | 0237
20 | 389
21 | 17
22 | 8
23 | 17
24 | 6
25 | 5

```

Figura 1.10: Diagrama de tallo y hojas para los datos de nicotina.

Tabla 1.9: Datos para el ejemplo 1.6

Muestra	Mediciones	Muestra	Mediciones
1	29 36 39 34 34	16	35 30 35 29 37
2	29 29 28 32 31	17	40 31 38 35 31
3	34 34 39 38 37	18	35 36 30 33 32
4	35 37 33 38 41	19	35 34 35 30 36
5	30 29 31 38 29	20	35 35 31 38 36
6	34 31 37 39 36	21	32 36 36 32 36
7	30 35 33 40 36	22	36 37 32 34 34
8	28 28 31 34 30	23	29 34 33 37 35
9	32 36 38 38 35	24	36 36 35 37 37
10	35 30 37 35 31	25	36 30 35 33 31
11	35 30 35 38 35	26	35 30 29 38 35
12	38 34 35 35 31	27	35 36 30 34 36
13	34 35 33 30 34	28	35 30 36 29 35
14	40 35 34 33 35	29	38 36 35 31 31
15	34 35 38 35 30	30	30 34 40 28 30

Existen otras formas en las que las gráficas de caja y bigote, y otras presentaciones gráficas, pueden ayudar al analista. Las muestras múltiples se pueden comparar de forma gráfica. Los diagramas de los datos pueden sugerir relaciones entre las variables y las gráficas ayudan a detectar anomalías u observaciones extremas en las muestras.

Existen otros tipos diferentes de diagramas y herramientas gráficas, los cuales se estudiarán en el capítulo 8 después de presentar otros detalles teóricos.

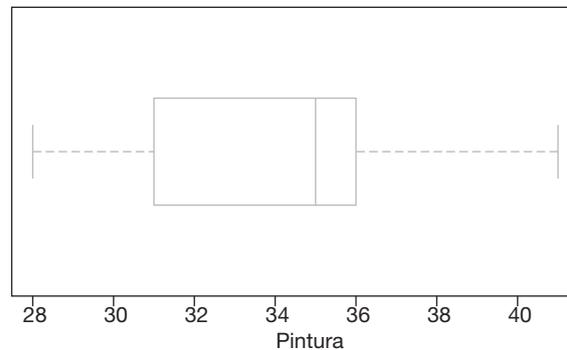


Figura 1.11: Gráfica de caja y bigote del grosor de las “asas” de latas de pintura.

Otras características distintivas de una muestra

Hay características de la distribución o de la muestra, además de las medidas del centro de localización y variabilidad, que definen aún más su naturaleza. Por ejemplo, en tanto que la mediana divide los datos (o su distribución) en dos partes, existen otras medidas que dividen partes o segmentos de la distribución que pueden ser muy útiles. Una separación en cuatro partes se hace mediante *cuartiles*, donde el tercer cuartil separa el cuarto (25%) superior del resto de los datos, el segundo cuartil es la mediana y el primer cuartil separa el cuarto (25%) inferior del resto de los datos. La distribución puede dividirse incluso más detalladamente calculando los percentiles. Tales cantidades dan al analista una noción de las denominadas *colas* de la distribución (es decir, los valores que son relativamente extremos, ya sean pequeños o grandes). Por ejemplo, el percentil 95 separa el 5% superior del 95% inferior. Para los extremos en la parte inferior o *cola inferior* de la distribución prevalecen definiciones similares. El primer percentil separa el 1% inferior del resto de la distribución. El concepto de percentiles desempeñará un papel significativo en buena parte de lo que estudiaremos en los siguientes capítulos.

1.7 Tipos generales de estudios estadísticos: diseño experimental, estudio observacional y estudio retrospectivo

En las siguientes secciones destacaremos el concepto de muestreo de una población y el uso de los métodos estadísticos para aprender o quizá para reafirmar la información relevante acerca de una población. La información que se busca y que se obtiene mediante el uso de tales métodos estadísticos a menudo influye en la toma de decisiones, así como en la resolución de problemas en diversas áreas importantes de ingeniería y científicas. Como ilustración, el ejemplo 1.3 describe un experimento sencillo, en el cual los resultados brindan ayuda para determinar los tipos de condiciones en los que no se recomienda utilizar una aleación de aluminio específica que podría ser muy vulnerable a la corrosión. Los resultados serían útiles no sólo para quienes fabrican la aleación, sino también para los clientes que consideren adquirirla. Este caso, y muchos otros que se incluyen en los capítulos 13 a 15, resaltan el concepto de condiciones experimentales diseñadas o controladas (combinaciones de condiciones de recubrimiento y humedad), que son de interés para aprender sobre algunas características o mediciones (nivel de corrosión) que