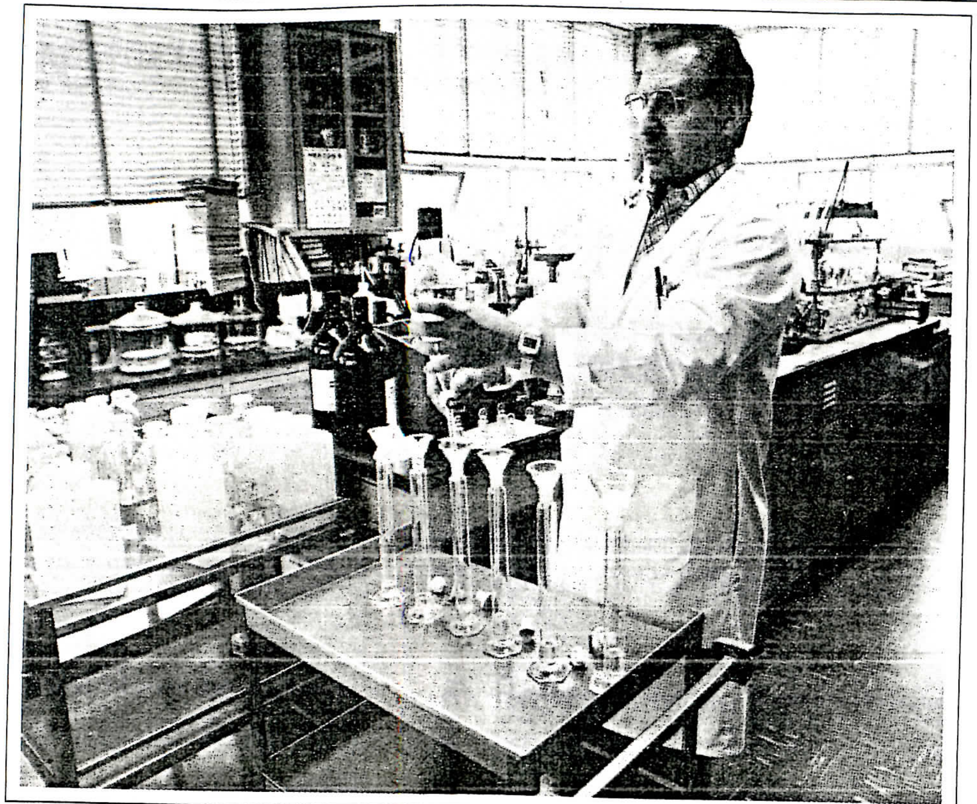
Segunda Edición

ESTADÍSTICA PARA ADMINISTRADORES

Richard I. Levin

2 Organización de los datos para que transmitan un significado: tablas y gráficas

OBJETIVOS: En los capítulos 2, 3 y 4 se expondrán los conceptos y técnicas de la estadística descriptiva. En el capítulo 2 se examinan dos métodos con los cuales se describe un conjunto de datos: tablas y gráficas. Si alguna vez ha oído un largo y tedioso informe sobre las cuotas que deben todavía ocho miembros de su club y ha deseado tener una representación gráfica para aliviar el tedio, seguramente ya está en condiciones de entender lo que se verá en el capítulo 2.



El ingeniero del control de calidad del agua en Charlotte (North Carolina) es responsable del nivel de clorinación del agua. Dicho nivel ha de acercarse bastante al que exige el departamento de salubridad. Para vigilar el cloro sin necesidad de verificar cada galón de agua que sale de la planta, el ingeniero muestrea diariamente algunos galones, mide el contenido de cloro y extrae una conclusión sobre el nivel promedio de clorinación que tiene el agua tratada ese día. La tabla anexa muestra las concentraciones de cloro de 30 galones seleccionados como muestra de un día. Estos niveles son los datos brutos de donde el ingeniero saca sus conclusiones respecto a la población total a la que se aplicó la clorinación ese día.

Concentraciones de cloro en partes por millón (ppm)
en 30 galones de agua tratada

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

Aplicando los métodos que se describen en este capítulo, podremos ayudar a este ingeniero a llegar a las conclusiones correctas.

Algunas definiciones

Los *datos* son colecciones de un número cualquiera de observaciones relacionadas entre sí. Podemos reunir la cantidad de teléfonos que varios trabajadores instalan en un día determinado o que uno de ellos instala por día durante un periodo de varios días; a los resultados podemos llamarlos *datos*. Una colección de datos recibe el nombre de *conjunto de datos*, y se da el nombre de *punto* (*observación*) *de datos* a una sola observación.

2-1 ¿COMO PODEMOS ARREGLAR LOS DATOS?

Para que los datos sean útiles, hemos de organizar nuestras observaciones de manera que podamos seleccionar tendencias y llegar a conclusiones lógicas. En este capítulo describiremos las técnicas que nos permiten organizar los datos en forma tabular y gráfica. En el capítulo 3 veremos cómo emplear los números para describir los datos.

Colección de datos

Representar a todos los grupos

Los estadísticos seleccionan sus observaciones de manera que todos los grupos relevantes estén representados en los datos. Para determinar el mercado potencial de un nuevo producto, los analistas podrían estudiar a 100 consumidores que viven en cierta zona geográfica. Y deben cerciorarse de que el grupo contenga una gran diversidad de personas que representen variables como nivel de ingresos, raza, escolaridad y barrio.

Obtener datos de observaciones o registros

Los datos provienen de observaciones reales o de documentos que se conservan para usos ordinarios. Por ejemplo, los hospitales tienen expedientes con el número de pacientes que usan los servicios de rayos X, pues esta información ayuda a preparar los informes de los médicos y a hacer la facturación. Sin embargo, esa información también puede organizarse para producir datos que el estadístico describe e interpreta.

Utilizar los datos referentes al pasado para tomar decisiones sobre el futuro

Los datos ayudan a los encargados de la toma de decisiones a hacer conjeturas bien fundamentadas acerca de las *causas* y, por tanto, sobre los *efectos* probables de ciertas características en algunas situaciones. Por lo demás, el conocimiento de las tendencias adquirido con la experiencia permite conocer los posibles resultados y planear con anticipación. Nuestra investigación del mercado quizá revele que el producto es preferido por las mujeres negras de las comunidades suburbanas, por los consumidores que perciben ingresos promedio y tienen una escolaridad también promedio. El mensaje publicitario del producto deberá dirigirse a esta audiencia meta. Y si los expedientes del hospital indican que más pacientes usaron los servicios de rayos X en junio que en enero, el departamento de personal habrá de determinar si ese fenómeno se dio sólo en el presente año o si es una indicación de una tendencia constante. De ser así, habrá de ajustar su programa de contrataciones y vacaciones para tenerla en cuenta.

Cuando los datos se organizan en forma compacta y útil, los encargados de la toma de decisiones consiguen información confiable del ambiente y se valen de ella para llegar a decisiones inteligentes. En el momento actual, las computadoras permiten a los estadísticos reunir enormes volúmenes de observaciones y condensarlas instantáneamente en tablas, gráficas y números. Se trata de formas compactas y utilizables, pero cabe preguntar: ¿son confiables? No se olvide que la exactitud de los datos que salen de la computadora depende de los que entran en ella. Como suelen decir los programadores de computadoras: "¡Entra basura, sale basura!" Los administradores deben tener mucho cuidado y asegurarse de que los datos que están utilizando se basan en suposiciones e interpretaciones correctas. Para poder confiar en la interpretación de unos datos cualesquiera, antes se prueban formulando las siguientes preguntas:

Pruebas de datos

1. ¿De dónde proceden los datos? ¿Está prejuiciada la fuente; es decir, tiende a tener interés en suministrar puntos de datos que lleven a una conclusión y no a otra?
2. ¿Apoyan o contradicen los datos la otra evidencia con que contamos?
3. ¿Existen datos que ignoramos y que nos harían llegar a una conclusión diferente?
4. ¿Cuántas observaciones tenemos? ¿Representan a todos los grupos que queremos estudiar?
5. ¿Es lógica la conclusión? ¿Hemos sacado conclusiones que los datos no apoyan?

Estudie sus respuestas a las preguntas anteriores. ¿Vale la pena utilizar los datos? ¿O más bien conviene esperar y recabar más información antes de actuar? Si en el hospital hiciera falta personal porque se contrató a muy pocas enfermeras para la sala de rayos X, ello se debería a que la administración se basó en datos insuficientes. Si la agencia publicitaria dirigiera su mensaje promocional exclusivamente a las amas de casa que viven en zonas suburbanas mientras podría haber triplicado sus ventas dirigiéndolo también a las que habitan en los suburbios de blancos, ello indicará que se basó en datos insuficientes. En uno y otro caso, probar los datos disponibles habría ayudado a los administradores a tomar decisiones más adecuadas.

Ejemplo de conteo doble

El efecto de datos incompletos o prejuiciados puede ilustrarse con el siguiente ejemplo. Una asociación de transporte proclamó en un anuncio que "el 75% de todo lo que consumimos es transportado por camión". Tal aserción podría hacernos creer que los automóviles, el ferrocarril, los aviones, los barcos y otros medios de transporte trasladan apenas 25% de lo que usamos. Llegar a semejante conclusión es fácil, mas no esclarecedor. En la afirmación de las compañías de transporte no figura la cuestión del "conteo doble". ¿Qué criterio aplicaban cuando algo era traído a nuestra ciudad por ferrocarril y entregado a nuestro domicilio por camión? ¿O cómo hicieron sus cálculos de los paquetes cuando fueron enviados por correo aéreo y luego entregados por motocicleta? Una vez resuelta la cuestión del conteo doble (muy difícil de resolver, por cierto), se descubre que la proporción de los productos de consumo transportados por camión es menor que la que afirman las compañías camioneras. Aunque este medio ayuda a *entregar* una proporción relativamente alta de lo que usamos, el ferrocarril y el barco transportan más mercancía para un mayor total de kilómetros.

Diferencia entre muestras y poblaciones

Definición de muestra y de población

Los estadísticos extraen datos de las muestras. Y esta información les sirve para hacer inferencias sobre la población que la muestra representa. Así pues, *muestra* y *población* son términos relativos. La población es un todo, y la muestra es una parte o segmento de él.

Función de las muestras

Estudiaremos las muestras a fin de poder describir las poblaciones. Nuestro hospital puede estudiar un pequeño grupo representativo de expedientes de rayos X en vez de examinar cada expediente de los últimos quince años. La agencia de encuestas Gallup puede entrevistar una muestra de apenas 2,500 adultos norteamericanos, a fin de predecir la opinión de todos los adultos que viven en Estados Unidos. Evidentemente es más fácil estudiar muestras que poblaciones enteras, y es un método confiable si se aplica con rigor y debidamente.

Función de las poblaciones

La *población* es una colección de todos los elementos que estamos estudiando y acerca de los cuales intentamos extraer conclusiones. Debemos definirla, de manera que quede claro si un elemento es o no miembro de ella. La población de nuestro estudio de mercado puede estar constituida por todas las mujeres que habitan en un radio de 15 km del centro de la ciudad de Cincinnati, que perciben ingresos anuales familiares entre $10,000 y $25,000 y que han cursado por lo menos once años de educación formal. Una mujer que viva en el centro de Cincinnati, que tenga un ingreso familiar de $15,000 y un grado universitario formará parte de esta población. Una mujer que viva en San Francisco, cuyo ingreso familiar sea de $7,000 o que haya cursado cinco años de educación formal no podrá figurar en esta población.

Necesidad de una muestra representativa

La *muestra* es una colección de algunos de los elementos que componen una población. La población de nuestro estudio de mercado la constituyen *todas* las mujeres que satisfacen los requisitos que acabamos de señalar. Cualquier grupo de mujeres que cumplan con ellos puede ser una muestra, a condición de que sea sólo una parte de la población entera. Una gran porción de un pastel de cerezas con unos trozos de corteza es una muestra de pastel; pero no es una muestra representativa, puesto que la proporción de los ingredientes no es la misma en la muestra que en el todo.

La *muestra representativa* contiene las características relevantes de la población en la *misma proporción* en que figuran en esa población. Si nuestra población de mujeres tiene una tercera parte de mujeres de raza negra, una muestra de ella que sea representativa respecto a la raza constará también de una tercera parte de mujeres de raza negra. En el capítulo 7 se explicarán a fondo algunos métodos concretos para hacer el muestreo.

Obtención de un patrón significativo en los datos

Los datos vienen en muchas formas

Hay muchas maneras de clasificar los datos. Podemos simplemente reunirlos y conservarlos en orden. Si las observaciones se miden en números, también podemos listar los puntos de datos por orden ascendente de valor numérico. Pero si los datos son trabajadores calificados (digamos carpinteros, albañiles o electricistas) que se necesitan en los sitios de construcción, si son diferentes tipos de automóviles fabricados por todas las empresas automotrices o si son los diversos colores de abrigos fabricados por determinada compañía, necesitaremos organizarlos de modo diferente. Tendremos que representar los puntos graficados (observaciones) de datos por orden alfabético o por algún otro principio organizador. Una forma útil de hacerlo consiste en dividir los datos en categorías o clases similares y luego contar el número de observaciones que caen dentro de cada categoría. Este método da origen a una *distribución de frecuencia* y se explica más adelante en el capítulo.

¿Por qué deben organizárse los datos?

La finalidad de organizar los datos es permitirnos ver rápidamente todas las características posibles de los datos que hemos recabado. Buscamos cosas como el intervalo (los valores mínimos y máximos), las tendencias notorias, aquello en torno a lo cual los datos tienden a agruparse, qué valores aparecen con mayor frecuencia y otros aspectos. Cuanto más abundante sea la información de este tipo que obtengamos de la muestra, mejor conoceremos la población de donde proviene y mejores serán las decisiones que tomemos.

EJERCICIOS

2-1 Cuando se les preguntó que tomarían si, en un naufragio, fueran arrojados a una isla desierta y tuvieran una sola opción de analgésicos, un número mayor de médicos prefirió Bayer a Tylenol, Bufferin o Advil. ¿Se extrajo esta conclusión de una muestra o de una población?

2-2 Uno de cada cinco automóviles vendidos en Estados Unidos durante 1986 fue fabricado en Japón. ¿Se extrajo esta conclusión de una muestra o de una población?

2-3 Una compañía de artículos electrónicos introdujo hace poco un nuevo amplificador y las pólizas de garantía indican que hasta ahora se han vendido 10,000 amplificadores. El presidente de la compañía, muy molesto después de leer 3 cartas de quejas sobre el nuevo amplificador, informó al gerente de producción que se aplicarían de inmediato costosas medidas de control de calidad para asegurarse de que no volvieran a aparecer los defectos. Comente la reacción del presidente desde el punto de vista de las 5 pruebas de datos que se describen en la página 10.

2-4 "El presidente no firmará este año el proyecto de ley agrícola", dijeron los líderes del partido demócrata del Congreso a principios de diciembre de 1985. Menos de 2 semanas después, el presidente estampó su firma en un proyecto de ley agrícola muy amplio. Mencione algunas causas de la predicción errónea de los líderes del partido demócrata.

2-5 Explique los datos incluidos en el problema de inicio de capítulo, basándose para ello en las 5 pruebas de datos.

2-2 EJEMPLOS DE DATOS BRUTOS

Se da el nombre *datos brutos* (*iniciales*) a la información antes de ser organizada y analizada. Se le llama así porque todavía no ha sido procesada por métodos estadísticos.

Problemas que afronta el personal de admisiones

Los datos de clorinación presentados en el problema con que principia el capítulo es un ejemplo de este tipo de datos. Examinemos otro ejemplo. Supongamos que el personal de admisiones de una universidad, preocupado por el éxito de los estudiantes a quienes selecciona, desea comparar sus calificaciones en la universidad con otros logros, como las calificaciones obtenidas en la enseñanza media, las puntuaciones de las pruebas y actividades extracurriculares. En lugar de estudiar a todos los estudiantes cada año, el personal de admisiones puede extraer una muestra de la población de todos los estudiantes en determinado periodo y examinar exclusivamente ese grupo, a fin de averiguar cuáles características parecen predecir el éxito. Por ejemplo, puede comparar las calificaciones logradas en la enseñanza media con las conseguidas en la universidad por cada integrante de la muestra. El personal puede asignarle un valor numérico a cada calificación. Después suma las calificaciones y las divide entre el número total de ellas para obtener el promedio de cada alumno. La tabla 2-1 contiene una muestra de estos datos brutos en forma tabular: 20 pares de calificaciones promedio en la enseñanza media y en la universidad.

Problema de la construcción de puentes

Cuando diseñan un puente, los ingenieros quieren conocer el esfuerzo que soportará un material determinado, el concreto por ejemplo. En vez de probar cada centímetro cuadrado para determinar su resistencia al esfuerzo, pueden tomar una muestra de concreto, probarla y sacar la conclusión del grado de esfuerzo que, en promedio, soportará ese tipo de concreto. La tabla 2-2 sintetiza los datos brutos reunidos de una muestra de 40 lotes de concreto que se empleará en la construcción de un puente.

TABLA 2-1 Promedio de calificaciones obtenidas por 20 estudiantes del último año en la enseñanza media y en la universidad

E. M.	UNIVERSIDAD	E. M.	UNIVERSIDAD	E. M.	UNIVERSIDAD	E. M.	UNIVERSIDAD
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

TABLA 2-2 Libras de presión por pulgada cuadrada que puede soportar 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

EJERCICIOS

2-6 Observe detenidamente los datos de la tabla 2-1. ¿Por qué esos datos requieren un arreglo ulterior? ¿Puede usted sacar algunas conclusiones de los datos tal como existen ahora?

2-7 El director de mercadotecnia de una gran empresa recibe un informe mensual sobre la actividad de ventas de uno de los productos de la empresa. El informe es un listado de las ventas del producto por estado durante el mes anterior. ¿Es éste un ejemplo de datos brutos?

2-8 El gerente de producción de una gran compañía recibe un informe mensual de la sección de control de calidad. El informe contiene la tasa de unidades rechazadas en la línea de producción (el número de unidades rechazadas por cada 100 que se producen), la máquina que causa el mayor número de ellas y el costo promedio de reparar las unidades rechazadas. ¿Es éste un ejemplo de datos brutos?

2-3 ORGANIZACION DE LOS DATOS MEDIANTE EL ARREGLO DE DATOS Y LA DISTRIBUCION DE FRECUENCIA

Definición del arreglo de datos

El arreglo de datos es una de las formas más sencillas de presentar los datos. Pone los valores en orden ascendente o descendente. La tabla 2-3 repite los datos de la clorinación incluidos en el problema con que empieza el capítulo, y en la tabla 2-4 se reorganizan esos números en un arreglo de datos por orden ascendente.

TABLA 2-3 Concentraciones de cloro en ppm de 30 galones de agua tratada

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 Arreglo de datos de las concentraciones de cloro en ppm de 30 galones de agua tratada

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

Ventajas de los arreglos de datos

Los arreglos de datos ofrecen varias ventajas sobre los datos brutos:

1. **Podemos descubrir rápidamente los valores mínimos y máximos en los datos.** En nuestro ejemplo de la clorinación, el intervalo es de 15.2 ppm a 16.9 ppm.

2. **Podemos dividir fácilmente los datos en secciones.** En la tabla 2-4, los primeros 15 valores (la mitad de la parte inferior de los datos) oscilan entre 15.2 y 16.0 ppm, y los últimos 15 valores (la mitad de la parte superior) fluctúan entre 16.0 y 16.9 ppm. De manera análoga, el tercio de la parte inferior de los valores oscila entre 15.2 y 15.8 ppm, el tercio de la mitad lo hace entre 15.9 y 16.2 ppm y el tercio de la parte superior lo hace entre 16.2 y 16.9 ppm.

3. **Podemos darnos cuenta si algunos valores aparecen más de una vez en el arreglo.** Los valores iguales aparecen juntos. En la tabla 2-4 se advierte que 9 niveles ocurrían más de una vez cuando se probó la muestra de 30 galones de agua.

4. **Podemos observar la distancia entre valores consecutivos de la tabla.** En la tabla 2-4, las concentraciones 16.6 y 16.8 son valores consecutivos. La distancia entre ellos es .2 ppm (16.8 − 16.6)

Desventajas de los arreglos de datos

Pese a las ventajas que acabamos de comentar, algunas veces un arreglo de datos no resulta práctico. Puesto que contiene todas las observaciones, es una forma engorrosa de resumir la información y de hacerla útil para la interpretación y la toma de decisiones. ¿Cómo puede hacerse eso?

Una mejor manera de arreglar los datos: la distribución de frecuencia

Con las distribuciones de frecuencia se manejan más datos

Una forma de sintetizar los datos consiste en valerse de una *tabla* o una *distribución de frecuencia*. Para captar la diferencia entre esto y el arreglo, tomaremos como ejemplo el inventario promedio (en días) de 20 tiendas de conveniencia.

En las tablas 2-5 y 2-6 hemos incluido datos idénticos referentes al inventario promedio y los hemos dispuesto primero como un arreglo en orden ascendente y luego como una distribución de frecuencia. Para obtener la tabla 2-6, tuvimos que dividir los datos en grupos de valores semejantes. A continuación registramos el número de puntos graficados (observaciones) de datos que caían dentro de cada grupo. Nótese que perdimos un poco de información al construir la distribución de

Se pierde un poco de información

TABLA 2-5 Arreglo de datos del inventario promedio (en días) de 20 tiendas de artículos de conveniencia

2.0	3.4	3.8	4.1	4.1	4.3	4.7	4.9	5.5	5.5
3.4	3.8	4.0	4.1	4.2	4.7	4.8	4.9	5.5	5.5

TABLA 2-6 Distribución de frecuencia del inventario promedio (en días) de 20 tiendas de artículos de conveniencia (6 clases)

CLASE (GRUPO DE OBSERVACIONES DE DATOS CON VALORES SEMEJANTES)	FRECUENCIA (NUMERO DE OBSERVACIONES EN CADA CLASE)
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

Pero se gana otra información

frecuencia. Por ejemplo, ya no sabemos que el valor 5.5 aparece cuatro veces o que el valor 5.1 no aparece en absoluto. Pero, por otra parte, adquirimos información concerniente al *patrón* de los inventarios promedio. En la tabla 2-6 advertimos que el inventario promedio cae más a menudo en el intervalo de 3.8 a 4.3 días. Rara vez encontramos un inventario promedio en el intervalo comprendido entre 2.0 y 2.5 días o entre 2.6 y 3.1 días. Los inventarios en los intervalos entre 4.4 y 4.9 días y entre 5.0 y 5.5 días no predominan, pero ocurren con mayor frecuencia que algunos otros. De lo anterior se deduce que las distribuciones de frecuencia sacrifican algunos detalles, pero en cambio nos ofrecen nuevas perspectivas de los patrones de datos.

Función de las clases en una distribución de frecuencia

La distribución de frecuencia es una tabla que organiza 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 conveniencia. En la tabla 2-5, dicha característica tiene once valores distintos. Pero esos mismos datos podrían dividirse en cualquier número de clases. Así, la tabla 2-6 usa seis clases. Podríamos resumir aún más los datos y servirnos únicamente de dos clases "menor que 3.8" y "mayor o igual a 3.8". También podríamos aumentar el número de clases empleando intervalos más pequeños, como hicimos en la tabla 2-7.

¿Por qué se le llama distribución de "frecuencia"?

Una distribución de frecuencia muestra **el número de observaciones provenientes del conjunto de datos que caen dentro de cada una de las clases.** Si podemos determinar la frecuencia con que ocurren los valores en cada clase de un conjunto de datos, estaremos en condiciones de construir una distribución de frecuencia.

Características de las distribuciones de frecuencia relativa

Definición de la distribución de frecuencia relativa

Hasta ahora hemos expresado la frecuencia con que ocurren los valores en cada clase como el número total de observaciones de datos que caen dentro de dicha clase. También podemos expresar la frecuencia de cada valor como una *fracción* o *porcentaje* del número total de observaciones. La frecuencia de un inventario promedio, digamos de 4.4 a 4.9, es 5 en la tabla 2-6 y .25 en la tabla 2-8. Para obtener este último valor, hemos dividido la frecuencia de esa clase (5) entre el número total de observaciones en el conjunto de datos (20). La respuesta puede ex-

TABLA 2-7 Distribución de frecuencia del inventario promedio (en días) de 20 tiendas de artículos de conveniencia (12 clases)

CLASE	FRECUENCIA	CLASE	FRECUENCIA
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

presarse como una fracción (5/20), un decimal (.25) o un porcentaje (25%). Una *distribución de frecuencia relativa* presenta las frecuencias en fracciones o porcentajes.

TABLA 2-8 Distribución de frecuencia relativa de un inventario promedio (en días) para 20 tiendas de artículos de conveniencia

CLASE	FRECUENCIA	FRECUENCIA RELATIVA: FRACCIONES DE OBSERVACIONES EN CADA CLASE
De 2.0 a 2.5	1	.05
De 2.6 a 3.1	0	.00
De 3.2 a 3.7	2	.10
De 3.8 a 4.3	8	.40
De 4.4 a 4.9	5	.25
De 5.0 a 5.5	4	.20
	20	1.00 suma de las frecuencias relativas de todas las clases

Las clases son exhaustivas

Adviértase en la tabla 2-8 que la suma de todas las frecuencias relativas es igual a 1.00 o 100%. Esto sucede porque una distribución de frecuencia relativa parea cada clase con su fracción o porcentaje correspondiente de los datos totales. Por consiguiente, las clases en cualquier distribución de frecuencia simple o relativa son *exhaustivas*. Todos los datos encajan en una u otra categoría. Adviértase asimismo que las clases en la tabla 2-8 son *mutuamente excluyentes*; es decir, ninguna observación cae dentro de más de una categoría. La tabla 2-9 ilustra este concepto al comparar las clases mutuamente excluyentes con las que se traslapan. En las distribuciones de frecuencia no se da traslape entre las clases.

Las clases son mutuamente excluyentes

TABLA 2-9 Clases mutuamente excluyentes y que se traslapan

Mutuamente excluyentes	De 1 a 4	De 5 a 8	De 9 a 12	De 13 a 16
No mutuamente excluyentes	De 1 a 4	De 3 a 6	De 5 a 8	De 7 a 10

Clases de datos cualitativos

Hasta ahora, nuestras clases han estado constituidas por números y han descrito algún atributo cuantitativo de muestras de objetos. También podemos clasificar la información conforme a características cualitativas, como raza, religión y sexo, que no caen naturalmente en categorías numéricas. Igual que las clases de atributos cuantitativos, estas clases deben ser exhaustivas y mutuamente excluyentes. La tabla 2-10 muestra la manera de construir distribuciones de frecuencia simple y relativa usando el atributo cualitativo de las ocupaciones o profesiones.

Clases abiertas de listas que no son exhaustivas

La tabla 2-10 no contiene todas las ocupaciones a que se dedican los graduados de una universidad, pero aun así es exhaustiva. ¿Por qué? La clase "otros" abarca todas las observaciones que no encajan en una de las categorías enumeradas. Recurrimos a una palabra genérica como esta cada vez que la lista

TABLA 2-10 Ocupaciones de una muestra de 100 graduados del Colegio Central

CLASE OCUPACIONAL	DISTRIBUCION DE FRECUENCIA (1)	DISTRIBUCION DE FRECUENCIA RELATIVA (1) ÷ 100
Actor	5	.05
Banquero	8	.08
Hombre de negocios	22	.22
Químico	7	.07
Médico	10	.10
Agente de seguros	6	.06
Periodista	2	.02
Abogado	14	.14
Maestro	9	.09
Otras	17	.17
	100	1.00

no enumere específicamente todas las posibilidades. Por ejemplo, si nuestra característica puede ocurrir en cualquier mes del año, una lista completa incluirá doce categorías. Pero si deseamos listar únicamente ocho meses de enero a agosto, usaremos el término "otros" para incluir nuestras observaciones durante los cuatro meses de septiembre, octubre, noviembre y diciembre. Aunque nuestra lista no contiene explícitamente todas las posibilidades, es exhaustiva. La categoría "otros" se llama *clase abierta* cuando permite que sea ilimitado el extremo superior o inferior del esquema de clasificación. La última clase de la tabla 2-11 ("72 en adelante") es abierta.

Clases discretas

Los esquemas de clasificación pueden ser cuantitativos o cualitativos *y* discretos o continuos. Las clases *discretas* son entidades individuales que no pasan de una clase a la siguiente sin una ruptura. Son discretas las siguientes clases: el número de hijos de las familias, el número de camiones que poseen las compañías de mudanzas y las ocupaciones de los graduados universitarios. Los datos discretos son aquellos que pueden asumir sólo un número limitado de valores. Así,

TABLA 2-11 Edades de los residentes de un estado

CLASE: EDAD (1)	FRECUENCIA (2)	FRECUENCIA RELATIVA (2) ÷ 89,592
Del nacimiento 7	8,873	.0990
De 8 a 15	9,246	.1032
De 16 a 23	12,060	.1346
De 24 a 31	11,949	.1334
De 32 a 39	9,853	.1100
De 40 a 47	8,439	.0942
De 48 a 55	8,267	.0923
De 56 a 63	7,430	.0829
De 64 a 71	7,283	.0813
De 72 en adelante	6,192	.0691
	89,592	1.0000

los graduados de una universidad pueden clasificarse en médicos y químicos, pero no en una profesión intermedia. El precio de cierre de las acciones de AT&T puede ser 66¾ o en el equipo de baloncesto puede haber un jugador que mida 2⅒ m.

Clases continuas

Los datos *continuos* pueden pasar de una clase a la siguiente sin ruptura alguna. Contienen una medida numérica como el peso de unas latas de tomates, los kilogramos de presión sobre el concreto o el promedio de calificaciones de los universitarios del último año. Los datos continuos pueden expresarse en fracciones o en números enteros.

EJERCICIOS

2-9 Una compañía de transmisiones electrónicas registró como sigue el número de recibos de servicio prestado por cada una de sus 20 tiendas en el último mes:

808	641	628	731	641
446	342	545	909	568
335	459	727	848	649
229	347	309	575	757

La compañía piensa que una tienda realmente no puede esperar alcanzar financieramente el punto de equilibrio con menos de 450 servicios prestados mensualmente. Además su política es dar un bono financiero al gerente que genere más de 700 servicios al mes. Disponga los datos anteriores en un arreglo e indique cuántas tiendas no están consiguiendo el punto de equilibrio y cuántas ganan el bono.

2-10 Use los datos de la compañía de transmisiones eléctricas incluidos en el ejercicio anterior. El vicepresidente de la compañía ha establecido lo que llama "lista de vigilancia de tiendas", esto es, una lista de las tiendas cuya actividad de servicios es demasiado baja como para justificar una atención especial por parte de la oficina central. En esa categoría quedan incluidas las tiendas cuya actividad de servicios oscila entre 500 y 600 acciones al mes. ¿Cuántas tiendas debe haber en esa lista si nos basamos en la actividad correspondiente al último mes?

2-11 El número de horas que tardaron los mecánicos de transmisiones en quitar, reparar y reemplazar una transmisión en una de las tiendas de la compañía de transmisiones electrónicas se registran en un día de la última semana del modo siguiente:

2.9	3.4	5.4	3.6	2.7
4.4	5.4	3.2	4.6	3.3
2.3	3.3	6.7	2.2	4.4
5.5	3.3	6.7	8.7	4.1

Construya una distribución de frecuencia con intervalos de 1.0 horas a partir de estos datos. ¿A qué conclusión llegará respecto a la productividad de los mecánicos si se basa en esta distribución? Si la gerencia de la compañía piensa que más de 6.0 horas es una prueba de desempeño insatisfactorio, ¿existe un problema grave o menor en el desempeño de esta tienda en particular?

2-12 A continuación se transcriben las edades de 50 miembros de un programa de servicio social de un condado:

81	53	67	60	80	64	56	54	91	61
66	88	67	65	52	72	74	65	73	69
43	54	76	70	97	68	82	75	79	60
39	87	76	97	86	45	60	43	65	76
92	72	82	80	70	65	50	58	70	56

Con los datos anteriores construya las distribuciones de frecuencia relativa usando 7 y 13 intervalos iguales. Las políticas estatales de los programas de servicio social exigen que aproximadamente 40% de los participantes del programa sean mayores de 50 años.

a) ¿Se ajusta el programa a esa política?
b) ¿Le ayuda la distribución de frecuencia relativa con 13 intervalos a contestar mejor la parte (a) de la respuesta que la distribución con 7 intervalos?
c) Suponga que el director de los servicios sociales quiera conocer la proporción de participantes en el programa cuya edad fluctúa entre 45 y 80 años. ¿Podría estimar la respuesta con una distribución de frecuencia relativa que tenga 7 intervalos o con una que tenga 13?

2-13 Disponga los datos de la tabla 2-2 en la página 13 en un arreglo por orden descendente.
a) Suponga que la ley estatal requiere que el concreto de los puentes resista por lo menos 2,500 libras/pulgada cuadrada. ¿Cuántas muestras no pasarán esta prueba?
b) ¿Cuántas muestras soportarán una presión de por lo menos 2,497 libras/pulgada cuadrada, pero no resistirán una presión de más de 2,504 libras/pulgada cuadrada?
c) Al examinar el arreglo, deberá advertir que algunas muestras pueden soportar cantidades idénticas de presión. Lístelas junto con el número de muestras que puede soportar cada presión.

2-14 Con los datos de la tabla 2-1 en la página 13, disponga los datos en un arreglo descendente del promedio de calificaciones obtenidas en un plantel de enseñanza media. Después colóquelos en un arreglo ascendente. ¿Qué conclusiones sacará de los 2 arreglos que no hubiera podido obtener con los datos originales?

2-15 La Environmental Protection Agency seleccionó muestras de agua en 12 ríos y corrientes que desembocan en el Lago Erie. Las muestras fueron probadas en su laboratorio y clasificadas respecto al nivel de contaminantes sólidos suspendidos en cada una. Los resultados de las pruebas se dan en la siguiente tabla:

Muestra	1	2	3	4	5	6	7	8	9	10	11	12
Porcentaje de contaminación (en ppm)	35.4	45.3	67.3	57.4	52.9	32.1	41.2	50.7	60.8	47.3	38.6	46.2

a) Disponga los datos en un arreglo de los valores más altos a los más bajos.
b) Determine el número de muestras que tienen un contenido de contaminación que oscile entre 30.0 y 39.9, 40.0 y 49.9, 50.0 y 59.9, 60.0 y 69.0.
c) Si 45.0 es el número empleado por la Environmental Protection Agency para indicar una contaminación excesiva, ¿cuántas muestras serán clasificadas en el grupo de las que contienen una contaminación excesiva?

2-16 Suponga que el personal de admisiones mencionado al describir la tabla 2-1 en la página 13 desea examinar la relación existente entre la diferencia del examen de la prueba de aptitudes académicas (SAT) (o sea la que hay entre la calificación real y la prevista basada en el promedio de calificaciones conseguidas por el alumno en la enseñanza media) y la dispersión del promedio de las calificaciones obtenidas en la enseñanza media y en la universidad (la diferencia entre unas y otras). El personal de admisión se servirá de los siguientes datos:

PROM. DE CALIF. EN ENS. MEDIA	PROM. DE CALIF. EN UNIV.	PUNTUACION EN SAT.	PROM. DE CALIF. EN ENS. MEDIA	PROM. DE CALIF. EN UNIV.	PUNTUACION EN SAT.
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 ha recibido la siguiente información del Educational Testing Service:

PROM. DE CALIF. EN ENS. MEDIA	PUNTUACION PROMEDIO EN SAT.	PROM. DE CALIF. EN ENS. MEDIA	PUNTUACION PROM. EN SAT.
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
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		

a) Disponga los datos anteriores en un arreglo que abarque desde los valores más altos hasta los más bajos. (Considere positivo un aumento en las calificaciones de la universidad respecto a las conseguidas en la enseñanza media y considere negativa una disminución en las calificaciones de la universidad que sean menores a las logradas en la enseñanza media.) Incluya en cada dispersión el diferencial correspondiente de la prueba de aptitudes académicas (SAT). (Considere negativa una calificación en dicha prueba si está por debajo de lo esperado y positiva si está por arriba de lo esperado.)
b) ¿Cuál es la dispersión más común?
c) Para la dispersión en la parte (b), ¿cuál es el diferencial más común de la prueba de aptitudes académicas?
d) ¿Qué conclusión puede extraer del análisis que ha hecho?

2-4 CONSTRUCCION DE UNA DISTRIBUCION DE FRECUENCIA

Ahora que hemos aprendido a dividir una muestra en clases, ya estamos en condiciones de tomar datos brutos y construir realmente una distribución de frecuencia. Para resolver el problema de clorinación que viene en la primera página de este capítulo, seguiremos los tres pasos siguientes:

Clasificación de datos

1. **Escoger el tipo y número de clases para dividir los datos.** En este caso, ya hemos optado por clasificar los datos según la medida cuantitativa del número de ppm (partes por millón) del cloro en el agua tratada, en vez de hacerlo a partir de un atributo cualitativo como el color o el olor del agua. Después necesitamos decidir cuántas clases utilizar y el intervalo (la distancia) que debe comprender cada clase. El intervalo será dividido en clases *iguales*; es decir, el ancho del intervalo del inicio de una clase al inicio de la siguiente ha de ser idéntico en todas las clases. Si escogemos un ancho de .5 ppm para cada clase en el ejemplo de la clorinación, las clases serán las que aparecen en la tabla 2-12.

División del intervalo entre clases iguales

TABLA 2-12 Concentraciones de cloro en muestras de agua tratada con intervalos de clase de .5 ppm

CLASE EN PPM	FRECUENCIA
15.1-15.5	2
15.6-16.0	16
16.1-16.5	8
16.6-17.0	4
	30

Problemas de las clases desiguales

Si las clases no fuesen iguales y si el ancho de los intervalos difiriese entre ellas, tendríamos una distribución mucho más difícil de interpretar que una de intervalos iguales. Imaginemos cuán arduo sería interpretar los datos presentados en la tabla 2-13.

TABLA 2-13 Concentraciones de cloro en muestras de agua tratada usando intervalos desiguales de clase

CLASE	AMPLITUD DE LOS INTERVALOS DE CLASE	FRECUENCIA
15.1-15.5	15.6 − 15.1 = .5	2
15.6-15.8	15.9 − 15.6 = .3	8
15.9-16.1	16.2 − 15.9 = .3	9
16.2-16.5	16.6 − 16.2 = .4	7
16.6-16.9	17.0 − 16.6 = .4	4
		30

Usar de 6 a 15 clases

El número de clases depende de la cantidad de puntos graficados (observaciones) de datos y de la gama de los datos reunidos. Cuanto más sean las observaciones de datos o más amplia su gama, más clases se necesitarán para dividirlos. Desde luego, si tenemos sólo 10 puntos de datos, sería absurdo tener también 10 clases. Por lo regular, los estadísticos rara vez emplean menos de seis clases o más de quince.

Determinar el ancho de los intervalos de clase

Dado que necesitamos tomar los intervalos de clase de igual tamaño, el número de clases determina el ancho de cada una. Para calcular los intervalos, usaremos esta ecuación:

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

Debemos utilizar el *siguiente valor de las mismas unidades* porque estamos midiendo el *intervalo* entre el primer valor de una clase y el primer valor de la siguiente. En nuestro ejemplo de la clorinación, el último valor es 16.9 y, por tanto, 17.0 es el siguiente valor. Como estamos usando seis clases en nuestro ejemplo, el ancho de cada una será:

$$\frac{\text{Siguiente valor unitario después del mayor valor de los datos} - \text{Valor más pequeño de los datos}}{\text{Número total de los intervalos de clase}} \quad [2\text{-}1]$$

$$= \frac{17.0 - 15.2}{6}$$

$$= \frac{1.8}{6}$$

$$= .3 \text{ onzas} \leftarrow \text{amplitud de los intervalos de clase}$$

Examinar los resultados

Ya hemos terminado el paso 1. Hemos decidido clasificar los datos según la medida cuantitativa de cuantas ppm (partes por millón) se encuentran en el agua tratada. Escogimos seis clases para cubrir el intervalo de 15.2 a 16.9 y, en consecuencia, utilizaremos .3 ppm como el ancho de los intervalos de clase.

2. **Clasificar los puntos de datos en clases y contar el número de puntos en cada clase.** Esto lo hicimos en la tabla 2-14. Toda observación de datos encaja por lo menos en una clase y ninguna observación lo hace en más de una clase. Por consiguiente, nuestras clases son exhaustivas y mutuamente excluyentes. Obsérvese que el límite inferior de la primera clase corresponde a la menor observación de datos de la muestra, y que el límite superior de la última clase corresponde a la observación mayor de los datos.

3. **Mostrar las observaciones en una gráfica.** (Véase la Fig. 2-1.)

Estos tres pasos nos permiten disponer los datos en forma tabular y gráfica. En este caso, nuestra información se incluye en la tabla 2-14 y en la figura 2-1. Estas dos distribuciones de frecuencia omiten algunos de los detalles contenidos en los datos brutos de la tabla 2-3, pero facilitan captar las tendencias de los datos. Así, una característica patente es que la clase 15.8 − 16.0 contiene la mayor parte de los elementos; la clase 15.2 − 15.4 es la que contiene menos elementos.

Crear las clases y contar las frecuencias

TABLA 2-14 Concentraciones de cloro en muestras de agua tratada con intervalos de clase de .3 ppm

CLASE	FRECUENCIA
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	3
	30

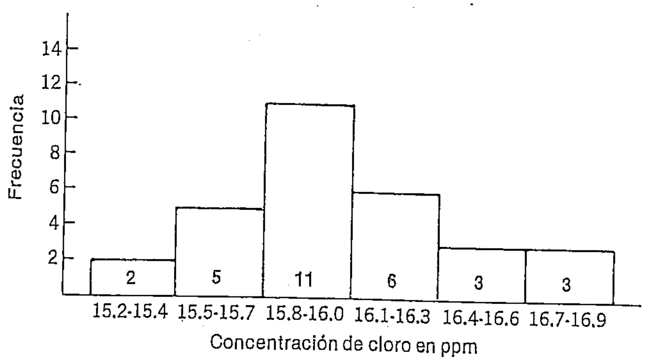


FIGURA 2-1 Distribución de frecuencia de las concentraciones de cloro en muestras de agua tratada usando intervalos de clase de .3 ppm

Captar las posibles tendencias

Obsérvese en la figura 2-1 que las frecuencias en las clases con un ancho de .3 ppm siguen una progresión regular: el número de puntos de datos comienza con 2 en la primera clase, aumenta a 5, llega a 11 en la tercera clase, disminuye a 6 y luego desciende hasta 3 en la quinta y sexta clases. Veremos luego que, cuanto más anchos sean los intervalos de clase, más gradual será la progresión. No obstante, si las clases son demasiado anchas, perderemos tanta información que la gráfica casi carece de significado. Por ejemplo, si reducimos la figura 2-1 a dos únicas categorías, obscureceremos la tendencia. Esto es evidente en la figura 2-2.

Uso de la computadora para construir distribuciones de frecuencia

Los cálculos manuales son engorrosos

A lo largo de este libro, recurriremos a ejemplos para mostrar cómo realizar muchas clases de análisis estadísticos. Por medio de esos ejemplos el lector sabrá qué clase de cálculos es preciso hacer. Confiamos que también logre entender los conceptos en que se basan éstos, de manera que entienda por qué son apropiados esos cálculos. Sin embargo, es un hecho que los cálculos manuales son engorro-

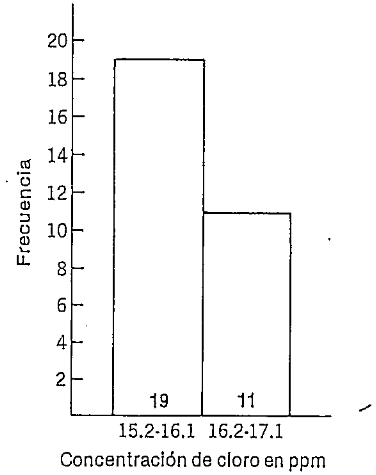


FIGURA 2-2 Distribución de frecuencia de las concentraciones de cloro en muestras de agua tratada usando intervalos de clase de 1 ppm

sos, fatigosos y muy expuestos a error. Algunos problemas reales tienen tantos datos que no es factible realizar los cálculos en forma manual.

Paquetes de software para el análisis estadístico

Por tal razón, en casi todo el mundo el análisis estadístico se realiza mediante computadoras. El usuario prepara los datos de entrada e interpreta los resultados del análisis tomando además las medidas pertinentes, pero la *máquina* lleva a cabo todos los "cálculos numéricos." Existen abundantes paquetes de software de uso generalizado con los cuales se hacen los análisis estadísticos: Minitab, SAS, SPSS y SYSTAT.* Nuestra intención es enseñarle al lector a utilizar cualquiera de ellos, pero aplicaremos el sistema SAS para ejemplificar las clases típicas de salida que producen dichos paquetes.

Uso de datos sobre calificaciones

El apéndice 9 contiene datos sobre las calificaciones de 199 estudiantes que emplearon este libro en nuestro curso de 1985. En la figura 2-3 hemos usado el programa SAS para crear una distribución de frecuencia de las puntuaciones brutas totales de los alumnos que asistieron al curso. A menudo también nos interesarán las *distribuciones de frecuencia bivariada*, en las cuales los datos se clasi-

CALIFICACIONES DE 1986 EN ESTADISTICA PARA ADMINISTRADORES SEGUN EL PROGRAMA SAS

TOTAL	FRECUENCIA	FRECUENCIA ACUM.	PORCENTAJE	PORCENTAJE ACUM.
[20-30)	1	1	0.503	0.503
[30-40)	1	2	0.503	1.005
[40-50)	9	11	4.523	5.528
[50-60)	27	38	13.568	19.095
[60-70)	68	106	34.171	53.266
[70-80)	65	171	32.663	85.930
[80-90)	26	197	13.065	98.995
[90-100)	2	199	1.005	100.000

FIGURA 2-3 Distribución de frecuencia del total de puntuaciones brutas

* Minitab es una marca registrada de Minitab, Inc., University Park, Pennsylvania. El programa SAS es una marca registrada del SAS Institute, Inc. Cary, N.C. El SPSS es una marca registrada de SPSS, Inc., Chicago, Illinois. El programa SYSTAT es una marca registrada de SYSTAT, Inc. Evanston, Illinois.

fican respecto a dos atributos distintos. En la figura 2-4 tenemos una distribución de ese tipo que muestra las calificaciones literales de las seis secciones del grupo escolar.

CALIFICACIONES DE 1988 EN
ESTADISTICA PARA ADMINISTRADORES
SEGUN EL PROGRAMA SAS

CALIFICACION SECCION

PORCENTAJE DE FRECUENCIA	1	2	3	4	5	6	TOTAL
F	2	3	0	1	3	2	11
	1.01	1.51	0.00	0.50	1.51	1.01	5.53
D	3	6	5	2	4	6	26
	1.51	3.02	2.51	1.01	2.01	3.02	13.07
C-	2	2	1	2	7	4	18
	1.01	1.01	0.50	1.01	3.52	2.01	9.05
C	9	11	3	9	6	6	44
	4.52	5.53	1.51	4.52	3.02	3.02	22.11
C+	3	6	10	6	7	2	34
	1.51	3.02	5.03	3.02	3.52	1.01	17.09
B-	1	5	5	1	0	3	15
	0.50	2.51	2.51	0.50	0.00	1.51	7.54
B	2	5	3	2	2	3	17
	1.01	2.51	1.51	1.01	1.01	1.51	8.54
B+	1	1	1	2	1	1	7
	0.50	0.50	0.50	1.01	0.50	0.50	3.52
A-	2	2	8	1	3	0	16
	1.01	1.01	4.02	0.50	1.51	0.00	8.04
A	2	5	1	0	3	0	11
	1.01	2.51	0.50	0.00	1.51	0.00	5.53
TOTAL	27	46	37	26	36	27	199
	13.57	23.12	18.59	13.07	18.09	13.57	100.00

FIGURA 2-4 Distribución de frecuencia bivariada que muestra las calificaciones en cada sección

EJERCICIOS

2-17 High Performance Bycicle Products Company, una empresa situada en Chapel Hill (North Carolina), muestreó sus registros de embarque durante cierto día, obteniendo los siguientes resultados:

TIEMPO TRANSCURRIDO DESDE LA RECEPCION
DE LA ORDEN HASTA LA ENTREGA (EN DIAS)

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

Construya una distribución de frecuencia para los datos anteriores y una distribución de frecuencia relativa. Use intervalos de 6 días.

a) ¿Qué afirmación puede hacer sobre la eficacia del procesamiento de pedidos a partir de la distribución de frecuencia?

b) Si la compañía quiere asegurarse de que la mitad de sus entregas se realicen en 10 o menos días, ¿puede usted determinar mediante la distribución de frecuencia si la compañía ha alcanzado su meta?

c) ¿Una distribución de frecuencia relativa le permite hacer con los datos una cosa que sería difícil efectuar con una simple distribución de frecuencia?

2-18 Use la tabla 2-2 de la página 13 para construir una distribución de frecuencia relativa empleando intervalos de 4.0 libras/pulgada cuadrada. ¿Qué conclusiones saca de esta distribución?

2-19 El Bureau of Labor Statistics ha muestreado 30 comunidades en todo el país y ha explicado los precios en cada una de ellas al inicio y al final de agosto, a fin de averiguar aproximadamente cuánto ha cambiado en ese mes el índice de precios al consumidor. El cambio porcentual de precios en las 30 comunidades se da a continuación:

0.8	0.2	−0.1	0.1	−0.2	0.2	0.3	0.5	−0.1	−0.2
0.0	0.6	0.3	0.2	1.0	−0.4	0.0	0.1	0.3	0.1
−0.5	−0.2	0.0	0.4	0.6	0.0	0.1	0.2	0.1	0.3

a) Disponga los datos en un arreglo ascendente.

b) Con las siguientes clases de igual tamaño, cree una distribución de frecuencia y una distribución de frecuencia relativa: −0.5 a −0.2, −0.1 a 0.2, 0.3 a 0.6, 0.7 a 1.0.

c) ¿Cuántas comunidades tenían precios que no cambiaron o que si lo hicieron fue en menos de 1.0%?

d) ¿Son discretos o continuos los datos anteriores?

2-20 Sarah Anne Rapp, presidenta de Baggit Incorporated, acaba de conseguir datos brutos en un estudio de mercado realizado hace poco por su compañía. La encuesta se efectuó para medir la eficacia del nuevo eslogan de la compañía: "Cuando tienes prisa y deseas algo bueno y rápido, pide Baggit". Para determinar el efecto que el eslogan tiene en las ventas de Luncheon Baggits, a 20 personas se les preguntó cuántas cajas mensuales del producto habían comprado antes y después de que el eslogan se empleó en la campaña publicitaria. He aquí los resultados:

ANTES/DESPUES		ANTES/DESPUES		ANTES/DESPUES		ANTES/DESPUES	
3	6	2	1	4	6	9	10
4	3	5	10	2	6	1	2
2	7	7	6	5	7	3	2
4	5	7	9	7	4	4	9
6	6	3	5	3	4	1	1

a) Construya una distribución de frecuencia y una de frecuencia relativa con las respuestas "antes", usando para ello clases de 1 a 2, de 3 a 4, de 5 a 6, de 7 a 8 y de 9 a 10.

b) Resuelva la parte (a) para las respuestas "después".

c) Dé la principal razón por la cual conviene utilizar las mismas clases para las respuestas "antes" y "después".

d) Para cada par de respuestas "antes/después", reste la respuesta "antes" a la respuesta "después" a fin de obtener un número que llamaremos "cambio" (ejemplo: 6 − 3 = +3) y cree una distribución de frecuencia y una distribución de frecuencia relativa para "cambio" empleando las clases de −3 a −1, de 0 a 2 y de 3 a 5.

e) Basándose en la información reunida antes, señale si el nuevo eslogan ha o no ayudado a aumentar las ventas y dé una o dos razones de su conclusión.

2-21 A continuación se transcriben las edades de 30 personas que compraron grabadoras de video en Simphony Music Shop la semana anterior:

24	39	38	21	16	44	34	66	32	35
19	33	13	25	44	46	57	60	31	40
23	27	43	44	32	56	52	61	30	20

a) Con sólo observar los datos tal como están, ¿qué conclusiones puede extraer rápidamente sobre el mercado de la tienda Symphony?

b) Construya una clasificación cerrada de 6 categorías. ¿Qué conclusiones adicionales de mercado es posible sacar ahora?

2-22 Use los datos del ejercicio 2-21.

a) Construya una clasificación abierta de 5 categorías. ¿El hecho de contar con esa clasificación le permite llegar a alguna conclusión sobre el mercado de Symphony?

b) Ahora construya una distribución de frecuencia relativa que encaje en la clasificación abierta de 5 categorías. ¿El hecho de tener esta distribución le proporciona a Symphony información complementaria útil para su estudio de mercadotecnia? ¿Por qué?

2-23 John Lyon, dueño de Fowler's Food Store en Chapel Hill (North Carolina), ha arreglado en la siguiente distribución de frecuencia las cantidades adquiridas por sus clientes la semana anterior:

$ GASTADOS	FRECUENCIA	$ GASTADOS	FRECUENCIA	$ GASTADOS	FRECUENCIA
0.00– 0.99	100	16.00–18.99	1,050	34.00–36.99	590
1.00– 3.99	250	19.00–21.99	990	37.00–39.99	400
4.00– 6.99	325	22.00–24.99	850	40.00–42.99	285
7.00– 9.99	450	25.00–27.99	800	43.00–45.99	110
10.00–12.99	925	28.00–30.99	760	46.00–48.99	80
13.00–15.99	1,000	31.00–33.99	700		

John dice que tener 17 intervalos definidos por 2 números es engorroso. ¿Puede usted ayudarle a simplificar los datos sin que pierdan demasiado de su valor?

2-24 A continuación se transcriben las marcas de clase (puntos medios de los intervalos) de una distribución que representa los minutos que tardaron los miembros del equipo de pista de una universidad en terminar un recorrido de cinco millas.

25 35 45

a) ¿Podría afirmar que el entrenador conseguirá suficiente información de esas marcas de clase para ayudarle a su equipo?

b) Si su respuesta a (a) es negativa, ¿cuántos intervalos le parecen idóneos?

2-25 El señor Mercado, un ingeniero de seguridad en una estación nuclear todos los días ha graficado la temperatura máxima del reactor durante el año pasado y preparó la siguiente distribución de frecuencia:

TEMPERATURA EN °C	FRECUENCIA
Menor a 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 encuentre en la distribución del señor Mercado.

2-26 Construya una clasificación cerrada discreta para las posibles respuestas a la parte relativa al "estado civil" de una solicitud de empleo. Construya también una clasificación abierta y discreta de 3 categorías para esas mismas respuestas.

2-27 Las cotizaciones de la bolsa de valores suelen contener el nombre de la compañía, las ofertas altas y bajas, el precio de las acciones al cierre y el cambio respecto al precio de cierre del día anterior. En seguida se da un ejemplo:

NOMBRE	POSTURA ALTA	POSTURA BAJA	CIERRE	CAMBIO
Sistemas Asociados	11½	10⅞	11¼	+½

Indique si una distribución de a) las acciones en la Bolsa de Valores de Nueva York por industria, b) precios al cierre en determinado día, c) cambios de precios en cierto día es
1) cuantitativa o cualitativa
2) continua o discreta
3) abierta o cerrada
¿Sería diferente su respuesta en (c) si el cambio fuera expresado simplemente como "más alto", "más bajo" o "inalterado"?

2-28 El nivel de ruido en decibeles de un avión que despega de un aeropuerto fue redondeado al decibel más próximo y agrupado en una distribución de frecuencia que tiene marcas de clase en 100 y 120. A menos de 100 decibeles el ruido no se considera fuerte en absoluto, mientras que si excede de los 130 resulta casi ensordecedor. Si los residentes de Villa Pacífica están reuniendo datos para entablar una demanda contra el aeropuerto, ¿es adecuada esta distribución para hacerlo? (Las marcas de clase son los puntos medios de los intervalos.)

2-29 Use los datos del ejercicio 2-28. Si el abogado que defiende al aeropuerto está recabando datos preparatorios para el juicio, aprobaría las marcas de clase del ejercicio 2-28 para su defensa?

2-30 El presidente de Ocean Airlines está tratando de estimar cuándo el Civil Aeronautics Board dará su dictamen sobre la solicitud de una nueva ruta para su empresa entre Charlotte y Nashville. Sus ayudantes han reunido los siguientes tiempos de espera de las solicitudes presentadas en el año anterior. Los datos se dan en días a contar desde la fecha de solicitud hasta el dictamen del Civil Aeronautics Board.

32	38	26	29	32	41	28	31	45	36
45	35	40	30	31	40	27	33	28	30
30	41	39	38	33	35	31	36	37	32
23	45	39	37	38	36	33	35	42	38
34	22	37	43	52	32	35	30	46	36

a) Construya una distribución de frecuencia usando 10 intervalos cerrados, igualmente espaciados. ¿Qué intervalo ocurre con mayor frecuencia?

b) Construya una distribución de frecuencia usando 5 intervalos cerrados, igualmente espaciados. ¿Qué intervalo ocurre con mayor frecuencia?

c) Si el presidente de Ocean Airlines tuviera una distribución de frecuencia relativa para (a) o (b), ¿le ayudaría eso a estimar la respuesta que necesita?

2-31 Para hacer una evaluación del desempeño y ajustes en las cuotas, Rodolfo Trelles supervisó las ventas de automóviles por parte de sus 40 vendedores. En un periodo de un mes, vendieron los siguientes números de unidades:

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) Basándose en la frecuencia, ¿cuáles serán las marcas deseadas de clase (puntos medios de los intervalos)?
b) Construya una distribución de frecuencia y una distribución de frecuencia relativa que tengan el mayor número posible de marcas. Espacie uniformemente sus intervalos de modo que tengan por lo menos 2 automóviles de ancho.
c) Si las ventas menores que 7 unidades al mes se consideran un desempeño inaceptable, ¿cuál de las dos respuestas, (a) o (b), le ayuda más a identificar el grupo de vendedores insatisfactorios?

2-32 Una cadena de neverías intenta conservar en inventario sus 55 sabores de helado en cada una de sus tiendas. Su director de investigación de mercado afirma que llevar mejores registros en las tiendas es la clave para evitar que se agoten las existencias. Damián Martínez, director de operaciones de los establecimientos, reúne datos aproximándolos hasta el medio galón más cercano a la cantidad de cada sabor de helado que se vende diariamente. En un día nunca se usan más de 20 galones de un sabor.
a) ¿Es continua o discreta la clasificación de los sabores? ¿Cerrada o abierta?
b) ¿Es discreta o continua la clasificación por "cantidad de helado"? ¿Abierta o cerrada?
c) ¿Son cualitativos o cuantitativos los datos?
d) ¿Qué recomendaría que hiciera el señor Martínez para generar mejores datos destinados a la investigación de mercado?

2-33 Roberto Durán es el dueño y el que recoge los boletos en un transbordador que transporta personas y automóviles de Long Island a Connecticut. Cuenta con datos que indican el número de personas, y también el de automóviles, que han sido trasladados en el transbordador durante los dos últimos meses. Por ejemplo:

JULIO 3 NUMERO DE PERSONAS, 173 NUMERO DE AUTOS, 32

puede ser un dato típico que Roberto anota diariamente. Ha creado 6 clases de igual espaciamiento para registrar el número diario de personas, y las marcas de clase son 84.5, 104.5, 124.5, 144.5, 164.5 y 184.5. Las 6 clases del señor Durán para el número diario de automóviles tienen las 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 superior e inferior de las clases para el número de personas?
b) ¿Cuáles son los límites superior e inferior de las clases para el número de automóviles?

2-5 GRAFICACION DE LAS DISTRIBUCIONES DE FRECUENCIA

Identificación de los ejes horizontal y vertical

Las figuras 2-1 y 2-2 (en las páginas 24 y 25) son un preámbulo de lo que vamos a ver ahora: cómo presentar gráficamente las distribuciones de frecuencia. Las gráficas ofrecen datos en una representación bidimensional. Sobre el eje *horizontal*, mostramos los valores de la variable (la característica que estamos midiendo); por

ejemplo, la concentración de cloro en ppm (partes por millón). Sobre el eje *vertical*, marcamos las frecuencias de las clases mostradas sobre el eje horizontal. Así, la altura de las casillas en la figura 2-1 mide el número de observaciones en cada clase, marcadas sobre el eje horizontal.

Función de las gráficas

Las gráficas de distribución de frecuencia y de distribución de frecuencia relativa son útiles porque ponen de relieve y aclaran las tendencias que no se captan fácilmente en las tablas. Atraen la atención del lector sobre las tendencias de los datos. Las gráficas nos ayudan además a resolver los problemas concernientes a las distribuciones de frecuencia. Nos permiten estimar algunos valores con una simple ojeada y nos brindan una verificación gráfica de la veracidad de nuestras soluciones.

Histogramas

Descripción de los histogramas

Las figuras 2-1 y 2-2 (pp. 24, 25) son dos ejemplos de histogramas. El *histograma* es una serie de rectángulos, todos ellos de anchura proporcional a la gama de valores dentro de una clase y también de altura proporcional a los elementos que caen dentro de la clase. Si las clases que empleamos en la distribución de frecuencia tienen el mismo ancho, las barras verticales del histograma lo tendrán también. La altura de la barra de cada clase corresponde al número de elementos de esta última. En consecuencia, el área contenida dentro de cada rectángulo (ancho por altura) es el mismo porcentaje del área de todos los rectángulos que la frecuencia relativa de esa clase respecto a todas las observaciones realizadas.

Función de un histograma de frecuencias relativas

Un histograma que se sirve de la frecuencia relativa de las observaciones de datos en cada una de las clases y no del número real de observaciones recibe el nombre de *histograma de frecuencia relativa*. Este tiene la misma forma que un histograma de frecuencia absoluta hecho con el mismo conjunto de datos. Ello se debe a que, en ambos tipos de histograma, el tamaño relativo de cada rectángulo es la frecuencia de esa clase comparada con el número total de observaciones.

Recuérdese que la frecuencia relativa de una clase cualquiera es el número de observaciones en la clase dividido entre el número total de las que se hayan hecho. La suma de todas las frecuencias relativas para un conjunto de datos es igual a 1.0. Teniendo presente lo anterior, podemos convertir el histograma de la figura 2-1 en un histograma de frecuencia relativa como el que encontramos en la figura 2-3. Nótese que la única diferencia entre ambos es la escala vertical de la izquierda. La escala de la figura 2-1 es el número *absoluto* de observaciones en cada clase; en cambio, la escala de la figura 2-3 es el número de observaciones en cada clase como una *fracción* del número total de ellas.

Ventaja de la frecuencia relativa

Es útil poder presentar los datos en función de la frecuencia relativa de observaciones y no en función de la frecuencia absoluta en cada clase pues, aunque los números absolutos pueden cambiar (por ejemplo, al probar más galones de agua), la relación entre las clases permanece estable. El 20% de los galones puede caer dentro de la clase "16.1 – 16.3 ppm", sin importar si probamos 30 o 300 galones. Es fácil comparar los datos de diferentes tamaños de las muestras cuando recurrimos a los histogramas de frecuencia relativa.

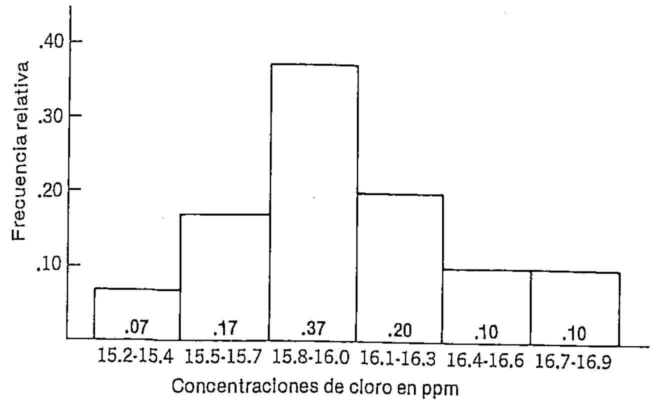


FIGURA 2-5 Distribución de frecuencia relativa de las concentraciones de cloro en muestras de agua tratada usando intervalos de clase de .3 ppm

Polígonos de frecuencias

Usar marcas de clase sobre el eje horizontal

Aunque de menor uso, los polígonos de frecuencias son otro medio de representar gráficamente tanto las distribuciones de frecuencia simple como las de frecuencia relativa. Para construir un polígono de frecuencias, marcamos las frecuencias sobre el eje vertical y los valores de la variable que vamos a medir las marcamos sobre el eje horizontal, tal como lo hicimos con los histogramas. El siguiente paso consiste en graficar cada frecuencia de clase dibujando un punto sobre su *marca de clase*, o punto medio, y en conectar los puntos consecutivos con una recta para formar un polígono (figura de muchos lados).

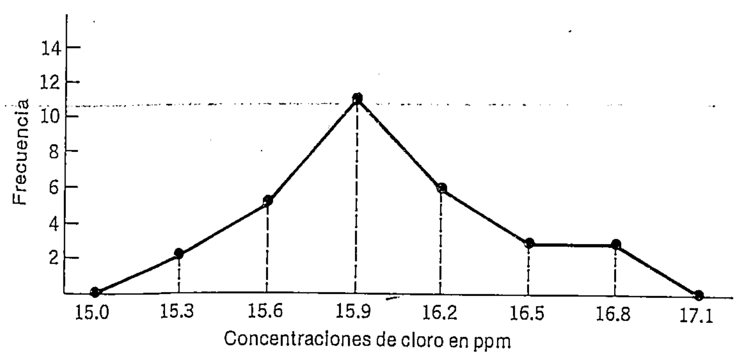


FIGURA 2-6 Polígono de frecuencias de las concentraciones de cloro en muestras de agua tratada utilizando intervalos de clase de .3 ppm

Agregar dos clases

La figura 2-6 es un polígono de frecuencias construido con los datos de la tabla 2-14 en la página 24. Si comparamos esta figura con la figura 2-1, advertiremos que se han agregado clases en *cada extremo* de la escala de los valores observados. Estas dos nuevas clases contienen cero observaciones, pero permiten al polígono alcanzar el eje horizontal en ambos extremos de la distribución.

Convertir un polígono de frecuencias en un histograma

¿Cómo se convierten un polígono de frecuencias en un histograma? El polígono es simplemente una gráfica lineal que une los puntos medios de todas las barras en un histograma. Por tanto, podemos reproducirlo trazando líneas verticales desde los límites de las clases (marcados sobre el eje horizontal) y uniéndolos con las líneas horizontales en las partes superiores del polígono en cada marca de clase. Esto lo hemos hecho con líneas punteadas en la figura 2-7.

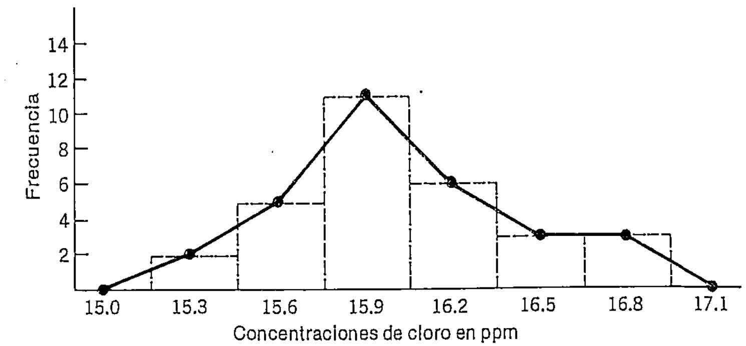


FIGURA 2-7 Histograma trazado de los puntos del polígono de frecuencias en la figura 2-6

Construcción de un polígono de frecuencias relativas

Se llama *polígono de frecuencias relativas* a aquel que usa la frecuencia relativa de los puntos de datos en cada clase y no el número real de puntos. Este polígono tiene la misma forma que el de frecuencia hecho con idéntico conjunto de datos pero con una diferente escala de valores sobre el eje vertical. En vez del número absoluto de observaciones, la escala es el número de las que hay en cada clase como una fracción de la cantidad total de ellas.

Ventajas del histograma

Los histogramas y los polígonos de frecuencia se parecen. ¿Por qué los necesitamos a ambos? He aquí las ventajas de los histogramas:

1. El rectángulo muestra claramente cada clase individual en la distribución.
2. El área de cada rectángulo, en relación con todos los otros, muestra la proporción del número total de observaciones que ocurren en esa clase.

Ventajas del polígono

Sin embargo, también los polígonos de frecuencias tienen algunas ventajas:

1. El polígono de frecuencias es más sencillo que su histograma correspondiente.
2. Ofrece un esquema más claro del patrón de datos.
3. El polígono se vuelve cada vez más suave y curvo, a medida que crece el número de clases y de observaciones.

Creación de una curva de frecuencia

Un polígono como el que acabamos de describir, suavizado por las clases y puntos de datos agregados, recibe el nombre de *curva de frecuencia*. En la figura 2-8 incluimos nuestro ejemplo de la clorinación, pero hemos aumentado el número de observaciones a 300 y a 10 el de las clases. Nótese que hemos unido los puntos con las líneas curvas para aproximar el aspecto que presentaría el polígono si tuviéramos un número infinito de observaciones de datos y clases de intervalo muy pequeñas.

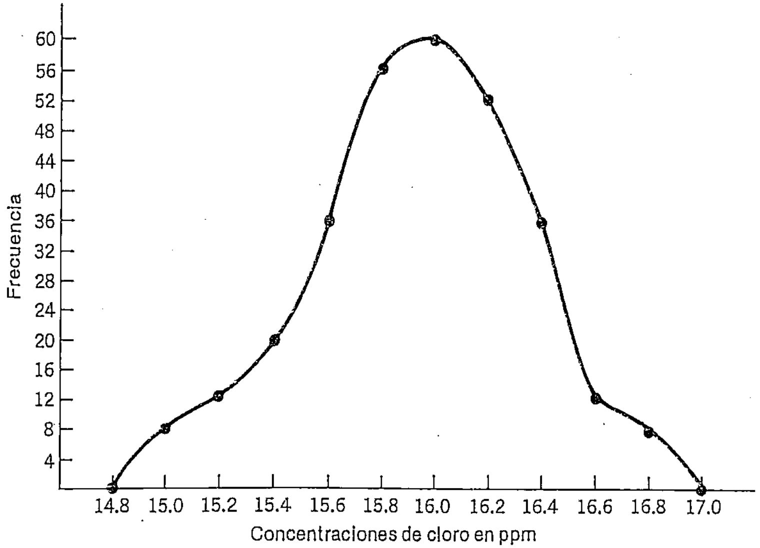


FIGURA 2-8 Curva de frecuencia de las concentraciones de cloro en 300 galones de agua, usando intervalos de .2 ppm

Ojivas

Definición de una distribución de frecuencia acumulativa

Una distribución de frecuencia acumulativa nos permite ver cuántas observaciones se hallan por arriba o debajo de ciertos valores, en lugar de limitarnos a anotar los números de elementos dentro de los intervalos. Por ejemplo, si queremos saber cuántos galones contienen menos de 17.0 ppm, podemos servirnos de una tabla que incluya las frecuencias acumulativas "menores que" en nuestra muestra, como se advierte en la tabla 2-15.

Tabla de frecuencias "menores que"

Una ojiva "menor que"

Se llama *ojiva* a la gráfica de una distribución de frecuencia acumulativa. La ojiva de una distribución de este tipo en la tabla 2-15 se muestra en la figura 2-9. Los puntos graficados representan la cantidad de galones que tienen menos cloro que las partes por millón indicadas sobre el eje horizontal. Nótese que el límite inferior de las clases en la tabla se convierte en el límite superior de la distribución de frecuencia acumulativa de la ojiva.

TABLA 2-15 Distribución de frecuencia acumulativa "menor que" de las concentraciones de cloro en ppm

CLASE	FRECUENCIA ACUMULATIVA
Menor que 15.2	0
Menor que 15.5	2
Menor que 15.8	7
Menor que 16.1	18
Menor que 16.4	24
Menor que 16.7	27
Menor que 17.0	30

En ocasiones la información que estamos utilizando se presenta a partir de frecuencias "mayores que". La ojiva apropiada para tal información tendrá una pendiente hacia abajo y hacia la derecha, no hacia arriba y hacia la izquierda como en la figura 2-9.

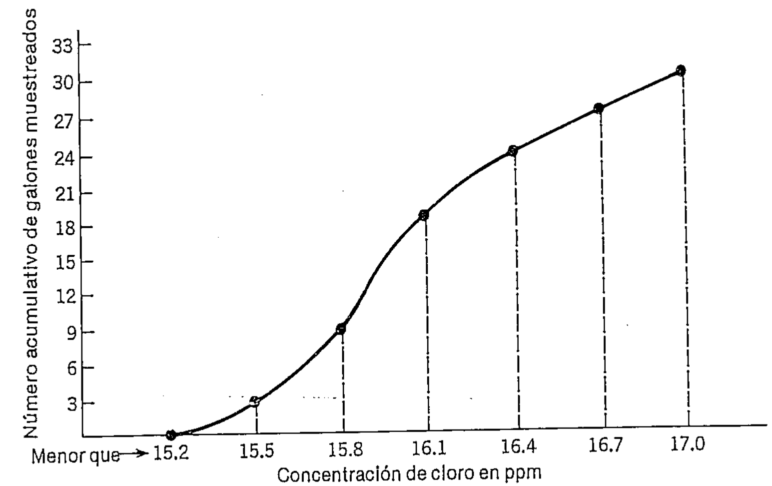


FIGURA 2-9 La ojiva "menor que" de la distribución de las concentraciones de cloro en ppm para 30 galones de agua tratada

Ojivas de frecuencias relativas

Podemos construir la ojiva de una distribución de frecuencia relativa de la misma manera que dibujamos la ojiva de una distribución de frecuencia absoluta en la figura 2-9. Habrá un cambio: la escala vertical. Igual que en la figura 2-5 de la página 32, esta escala debe marcar la *fracción* del número total de observaciones que caen dentro de cada clase.

TABLA 2-16 Distribución de frecuencia relativa acumulativa de las concentraciones de cloro en ppm

CLASE	FRECUENCIA ACUMULATIVA	FRECUENCIA ACUMULATIVA RELATIVA
Menor que 15.2	0	.00
Menor que 15.5	2	.07
Menor que 15.8	7	.23
Menor que 16.1	18	.60
Menor que 16.4	24	.80
Menor que 16.7	27	.90
Menor que 17.0	30	1.00

Para construir una ojiva de distribución de frecuencia acumulativa, podemos referirnos a la distribución de frecuencia relativa (como la Fig. 2-5) y elaborar una tabla que incluya los datos (por ejemplo, la Fig. 2-10). Nótese que las figuras 2-9 y 2-10 son equivalentes, exceptuado el eje vertical de la izquierda.

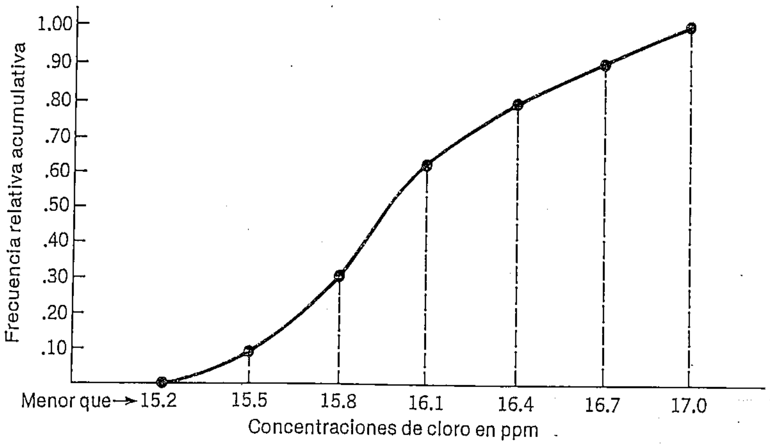


FIGURA 2-10 Ojiva "menor que" de la distribución de las concentraciones de cloro en ppm para 30 galones de agua tratada, usando frecuencias relativas

Aproximación al arreglo de datos

Supóngase que ahora trazamos una línea perpendicular al eje vertical en la marca .50 para intersectar nuestra ojiva. (Esto lo hicimos en la Fig. 2-11.) De este modo, podemos leer un valor aproximado de la concentración de cloro en el decimoquinto galón de un arreglo de 30 galones. Con ello, hemos retornado al

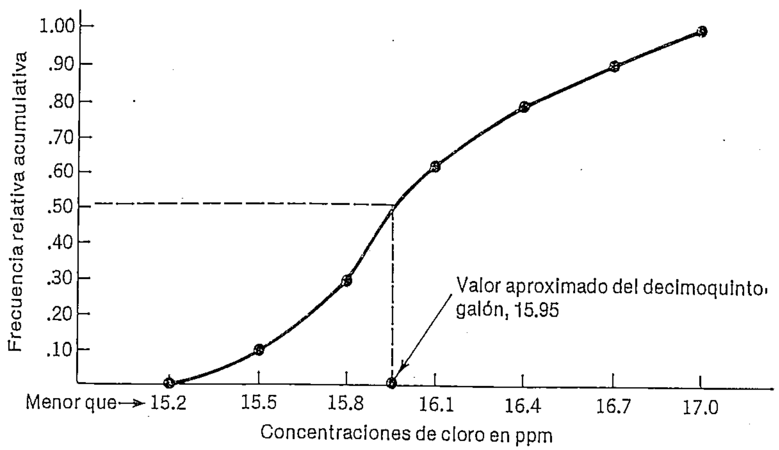


FIGURA 2-11 Ojiva "menor que" de la distribución de las concentraciones de cloro en ppm para 30 galones de agua tratada, lo cual indica el valor medio aproximado del arreglo original de datos

primer arreglo de datos explicado en el presente capítulo. Mediante ese arreglo podemos construir distribuciones de frecuencia. Con éstas podemos construir distribuciones de frecuencia acumulativa. Estas a su vez nos permiten graficar una ojiva. Y a partir de la ojiva estamos en condiciones de aproximar los valores que teníamos en el arreglo de datos. Sin embargo, normalmente es imposible recuperar los datos originales *exactos* partiendo de las representaciones gráficas que acabamos de mencionar.

Uso de la computadora para graficar las distribuciones de frecuencia

Uso del programa SAS para producir histogramas

Emplearemos el paquete estadístico SAS para producir algunos histogramas de los datos sobre las calificaciones que se dan en el apéndice 9. La figura 2-12 contiene un histograma de las calificaciones brutas totales de los alumnos. Nótese que este histograma tiene barras horizontales en vez de las barras verticales que hemos dibujado hasta ahora. Además, a la derecha de las barras, el programa SAS proporciona las frecuencias absolutas, las relativas y las acumulativas "menores que" (tanto absolutas como relativas).

En la figura 2-4 hemos observado una distribución de frecuencia bivariada. También podemos crear histogramas que contengan información sobre ambas variables. La figura 2-13 es un histograma vertical de las calificaciones de los alumnos, en el cual cada barra está dividida en dos segmentos que muestran las fracciones de

```
                    CALIFICACIONES DE 1985 EN
               ESTADISTICA PARA ADMINISTRADORES
                    SEGUN EL PROGRAMA SAS

                DIAGRAMA DE BARRAS DE FRECUENCIA
TOTAL DEL
PUNTO MEDIO
                                              FRECUENCIA  FREC.   PORCENTAJE PORCENTAJE
                                                          ACUMUL.            ACUMUL.

          |
    5     |                                     0          0       0.00       0.00
          |
   15     |                                     0          0       0.00       0.00
          |
   25     |*                                    1          1       0.50       0.50
          |
   35     |*                                    1          2       0.50       1.01
          |
   45     |*****                                9         11       4.52       5.53
          |
   55     |**************                      27         38      13.57      19.10
          |
   65     |**********************************  68        106      34.17      53.27
          |
   75     |*********************************   65        171      32.66      85.93
          |
   85     |*************                       26        197      13.07      98.99
          |
   95     |*                                    2        199       1.01     100.00
          |
      -----+----+----+----+----+----+-----
          10   20   30   40   50   60

                   FRECUENCIA
```

FIGURA 2-12 Histograma y distribuciones de frecuencia del total de las puntuaciones brutas los que reciben esa calificación y que estaban en las secciones a cargo de los profesores y de los estudiantes graduados que trabajaban como profesores asistentes (denotados mediante el sombreado de las barras con las P y las T).

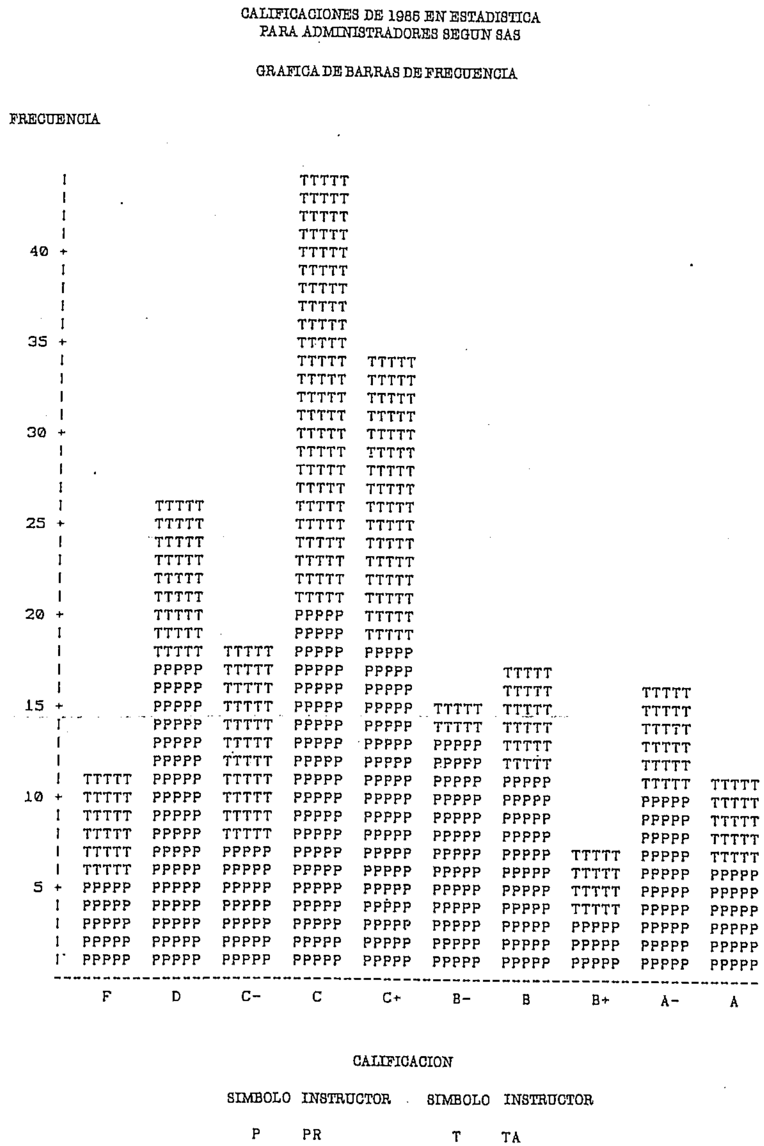


FIGURA 2-13
Histograma de las calificaciones que muestra el tipo de instructor

EJERCICIOS

2-34 A continuación se ofrece una distribución de frecuencia del peso de 150 personas que utilizaron un elevador cierto día. Construya un histograma con esos 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 observa en el histograma respecto a los datos que no captó de inmediato en la distribución de frecuencia?
b) Si cada elevador tiene capacidad para 6 personas pero su capacidad total de seguridad se limita a 400 kg, ¿qué puede hacer el operador para maximizar esa capacidad sin rebasar el límite de seguridad? ¿Apoyan los datos su propuesta?

2-35 A continuación se transcriben en pies los datos de las longitudes de una muestra de 25 botes que usan el North Canal en el Estado de Nueva York:

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 contestar las siguientes preguntas:
a) Se cobra una tarifa a los botes de más de 60 pies de largo. ¿Aproximadamente qué proporción de botes pasará el canal sin pagar la tarifa?
b) Se cobra una tarifa de 50¢ por pie a los botes de más de 60 pies de largo. ¿Aproximadamente cuántas veces recibirá el cobrador de las tarifas $12 o más?

2-36 Homer Willis, capitán de un barco pesquero de Salter Path (North Carolina) piensa que el punto de equilibrio de pesca para sus barcos es de 5,000 libras por viaje. A continuación se dan los datos de una muestra de pesca en 200 viajes que los barcos de Homer han hecho recientemente:

6500	6700	3400	3600	2000
7000	5600	4500	8000	5000
4600	8100	6500	9000	4200
4800	7000	7500	6000	5400

Construya una ojiva que le ayude a contestar las siguientes preguntas:
a) ¿Aproximadamente qué proporción de los viajes alcanzan el punto de equilibrio?
b) ¿Qué pesca representa el valor medio aproximado en el arreglo de datos de los barcos de Willis?
c) ¿Qué pesca de los barcos de Willis superan el punto de equilibrio un 80% de las veces?

2-37 Central Carolina Hospital tiene los siguientes datos que representan el peso neonatal en libras de 200 niños prematuros.

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 contestar las siguientes preguntas:

a) ¿Cuál fue el valor medio aproximado del conjunto original de datos?
b) Si normalmente a los niños prematuros de menos de 3.0 libras de peso se les mantiene como precaución en una incubadora varios días, ¿más o menos qué porcentaje de los prematuros que nacen en Central Carolina Hospital necesitarán una incubadora?

2-38 Antes de construir una presa sobre el Río Colorado, el Army Corps of Engineers efectuó una serie de pruebas para medir el flujo de agua más allá del sitio propuesto para la obra. Los resultados de las pruebas se usaron para preparar la siguiente distribución de frecuencia:

FLUJO DEL RIO (MILES DE GALONES POR MINUTO)	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

a) Con los datos de la tabla anterior construya una distribución y una ojiva de frecuencia acumulativa "mayor que".
b) Con los datos de la tabla anterior construya una distribución y una ojiva de frecuencia acumulativa "menor que".
c) Por medio de la ojiva estime qué proporción del flujo ocurre en menos de 1,300 galones por minuto.

2-39 Nora Velarde, asesora de una pequeña empresa de corretaje, estaba tratando de diseñar programas de inversión que fuesen atractivos para los ancianos. Sabía que, si los clientes potenciales podían ganar cierto nivel de rendimiento, estarían dispuestos a arriesgar su dinero; pero se mostrarían renuentes más allá de determinado nivel. De un grupo de 50 sujetos consiguió los siguientes datos acerca de los diversos niveles de rendimiento que se requería para que cada uno invirtiese $1,000:

PUNTO DE INDIFERENCIA	FRECUENCIA
$70–74	2
75–79	5
80–84	10
85–89	14
90–94	11
95–99	3
100–104	3
105–109	2
	50

a) Construya distribuciones de frecuencia acumulativa "mayores que" y "menores que".
b) Grafique las 2 distribuciones de la parte (a) convirtiéndolas en ojivas de frecuencia relativa.

2-40 En la oficina de un diario, el tiempo que se tardan en imprimir la primera plana fue registrado durante 50 días. A continuación se transcriben los datos, aproximados a décimas de minuto:

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) Disponga los datos en un arreglo por orden ascendente.
b) Construya con los datos una distribución de frecuencia y una distribución de frecuencia acumulativa "menor que", usando intervalos de .8 minutos.
c) Construya con los datos un polígono de frecuencias.
d) Construya con los datos una ojiva de frecuencia "menor que".
e) Por medio de la ojiva estime qué porcentaje de las veces la primera plana del periódico puede imprimirse en menos de 24 minutos.

2-41 Un agente de seguros tiene datos sobre la cantidad mensual de pólizas que vendió en los 3 últimos años. Sus datos los ha arreglado en la siguiente distribución de frecuencia.

VENTAS MENSUALES	FRECUENCIA
$1,000–$1,149	1
1,150– 1,299	3
1,300– 1,449	6
1,450– 1,599	4
1,600– 1,749	8
1,750– 1,899	9
1,900– 2,049	3
2,050– 2,199	2

a) Construya una distribución de frecuencia relativa.
b) Construya, sobre la misma gráfica, un histograma de frecuencia relativa y un polígono de frecuencias relativas.

2-42 La National Association of Real Estate Sellers ha recabado los siguientes datos en una muestra de 130 vendedores que representan los ingresos anuales totales por concepto de comisión:

UTILIDADES	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 contestar estas preguntas.

a) ¿Aproximadamente qué proporción de los vendedores gana más de $25,000?
b) ¿Aproximadamente cuánto gana el vendedor "medio" de la muestra?
c) ¿Aproximadamente cuánto podría ganar al año un agente de bienes raíces cuyo rendimiento fuera 25% del nivel máximo?

2-43 Un museo de historia natural cuenta con datos sobre el número de minutos que los visitantes ven cierta exhibición de un dinosaurio. He aquí esos datos, redondeados al minuto más cercano:

MINUTOS ANTE EL OBJETO EXHIBIDO	FRECUENCIA
Menor que 2	30
2– 3	40
4– 5	40
6– 7	90
8– 9	70
10–11	50
12–13	50
14–15	30
	400

a) Construya una distribución de frecuencia acumulativa "menor que".
b) Construya una ojiva basándose en la parte (a).
c) La administración ha decidido que un objeto de exhibición es un fracaso si 50% de los visitantes pasan menos de 4 minutos viéndolo. ¿Cuál es el porcentaje de los que lo observan menos de 4 minutos? Por otra parte, a la administración le gustaría saber aproximadamente cuántos minutos dedica a ver el objeto el visitante número 200, para poder tener un valor aproximado al respecto.

2-6 GLOSARIO DEL CAPITULO

ARREGLO DE DATOS Organización de los datos brutos por observaciones en orden ascendente o descendente.

CLASE ABIERTA Aquella que permite que sea ilimitado el extremo superior o inferior de un esquema de clasificación cuantitativa.

CONJUNTO DE DATOS Una colección de datos.

CURVA DE FRECUENCIA Un polígono de frecuencias suavizado al agregarle clases y observaciones de datos al conjunto de datos.

DATOS Colección de varias observaciones relacionadas en una o más variables.

DATOS BRUTOS Información antes de ser organizada o analizada mediante métodos estadísticos.

DATOS CONTINUOS Los que pueden pasar de una clase a la siguiente sin ruptura o que pueden expresarse como números enteros o fracciones.

DATOS DISCRETOS Los que no pasan de una clase a la siguiente sin ruptura; es decir, aquellos en que las clases representan categorías o conteos bien definidos y pueden ser representadas por números enteros.

DISTRIBUCION DE FRECUENCIA Representación organizada de los datos que muestra el número de observaciones del conjunto de datos que caen dentro de cada conjunto de clases mutuamente excluyentes.

DISTRIBUCION DE FRECUENCIA ACUMULATIVA Representación tabular de los datos que muestra cuántas observaciones se hallan encima o debajo de ciertos valores.

DISTRIBUCION DE FRECUENCIA RELATIVA Representación de datos que muestra la fracción o porcentaje del conjunto total de datos que caen dentro de cada conjunto de clases mutuamente excluyentes.

HISTOGRAMA Gráfica de un conjunto de datos que se compone de una serie de rectángulos, de anchura proporcional al intervalo de valores en una clase y de altura proporcional al número de observaciones que caen dentro de la clase o a la fracción de elementos de la clase.

MUESTRA Colección de algunos de los elementos de la población en estudio; se emplea para describir la población.

MUESTRA REPRESENTATIVA Aquella que contiene las características relevantes de la población en la misma proporción en que figuran en la población.

OJIVA Gráfica de una distribución de frecuencia acumulativa.

POBLACION Conjunto de todos los elementos que estamos estudiando y acerca de los cuales tratamos de sacar conclusiones.

POLIGONO DE FRECUENCIAS Gráfica lineal que une los puntos medios de cada clase en un conjunto de datos; se grafica en la altura correspondiente a la frecuencia de la clase.

PUNTO GRAFICADO (OBSERVACION) DE DATOS Observación individual de un conjunto de datos.

2-7 ECUACIONES UTILIZADAS EN EL CAPITULO

[2-1] $$\text{Amplitud de los intervalos de clase} = \frac{\text{Siguiente valor unitario después del valor máximo en datos} - \text{Valor mínimo en los datos}}{\text{Número total de intervalos de clase}}$$ p. 23

Para arreglar los datos escoja el número de clases en que los dividirá (normalmente entre 6 y 15) y luego aplique la ecuación 2-1 para determinar la *amplitud de los intervalos de clase de igual tamaño*. Esta fórmula usa el siguiente valor de las mismas unidades, porque mide el intervalo entre el primer valor de una clase y el primero de la siguiente clase.

2-8 EJERCICIOS DE REPASO

2-44 El siguiente conjunto de datos brutos da el nivel de ingresos y escolaridad en Estados Unidos de una muestra de individuos. ¿Podrá llegar a algunas conclusiones si los arregla? Rearréglelos en una forma que los haga más significativos.

INGRESOS	ESCOLARIDAD	INGRESOS	ESCOLARIDAD	INGRESOS	ESCOLARIDAD
$17,000	High school	$ 21,200	B.S.	$17,200	2 yrs. college
20,800	B.S.	28,000	B.S.	19,600	B.A.
27,000	M.A.	30,200	High school	36,200	M.S.
70,000	M.D.	22,400	2 yrs. college	14,400	1 yr. college
29,000	Ph.D.	100,000	M.D.	18,400	2 yrs. college
14,400	10th grade	76,000	Law degree	34,400	B.A.
19,000	High school	44,000	Ph.D.	26,000	High school
23,200	M.A.	17,600	11th grade	52,000	Law degree
30,400	High school	25,800	High school	64,000	Ph.D.
25,600	B.A.	20,200	1 yr. college	32,800	B.S.

2-45 Los 50 estados de la Unión Americana envían al ministerio del trabajo los siguientes datos: el número promedio de trabajadores que no asisten diariamente a sus labores en cada una de las 13 semanas de un trimestre financiero, así como el porcentaje de inasistencias en cada estado. ¿Es éste un ejemplo de datos brutos? Explique su respuesta.

2-46 El departamento de agricultura de Nebraska tiene los siguientes datos que representan el crecimiento semanal (en pulgadas) de muestras de maíz recién sembrado:

0.4	1.9	1.5	0.9	0.3	1.6	0.4	1.5	1.2	0.8
0.9	0.7	0.9	0.7	0.9	1.5	0.5	1.5	1.7	1.8

a) Disponga los datos en un arreglo por orden descendente.
b) Construya una distribución de frecuencia relativa usando intervalos de .25.
c) Con lo que ha hecho hasta ahora, ¿qué conclusiones puede sacar respecto al crecimiento en esta muestra?
d) Construya una ojiva que le ayude a determinar qué proporción del maíz creció más de 1.0 pulgadas a la semana.
e) ¿Aproximadamente cuál fue la tasa de crecimiento semanal del elemento medio de la muestra en el nuevo arreglo de datos?

2-47 El National Safety Council muestreó aleatoriamente la profundidad de la rodada de 60 llantas delanteras derechas de vehículos de pasajeros que se detenían en un área de descanso sobre una carretera interestatal. Con esos datos, construyeron la siguiente distribución de frecuencia:

PROFUNDIDAD DE LA RODADA (PULGADAS)	FRECUENCIA	PROFUNDIDAD DE LA RODADA (PULGADAS)	FRECUENCIA
16/32 (llanta nueva)	5	4/32–6/32	7
13/32–15/32	10	1/32–3/32	4
10/32–12/32	20	0/32 desgastada	2
7/32–9/32	12		

a) ¿Aproximadamente cuál fue la profundidad de la rodada de la trigésima llanta en el arreglo de datos?
b) Si la profundidad de la rodada menor que 7/32 se considera peligrosa, ¿aproximadamente qué proporción de las llantas en la carretera son poco seguras?

2-48 Una compañía fabrica 15 productos básicos. La compañía conserva registros del número de cada producto fabricado por mes, a fin de examinar los niveles relativos de producción. Los registros muestran que los siguientes números de cada producto fueron fabricados por ella en el último mes de 20 días laborales:

9,897	10,052	10,028	9,722	9,908
10,098	10,587	9,872	9,956	9,928
10,123	10,507	9,910	9,992	10,237

Construya una ojiva que le ayude a contestar las siguientes preguntas:
a) ¿En cuántos de sus artículos la producción rebasó el punto de equilibrio de 10,000 unidades?
b) ¿Qué nivel de producción rebasó el 75% de sus productos en el mes?
c) ¿Qué nivel de producción alcanzó el 90% de sus productos en ese mes?

2-49 El administrador de un hospital ha ordenado un estudio sobre el tiempo que un paciente espera antes de ser atendido por el personal de la sala de urgencias. Los siguientes datos fueron reunidos durante una jornada típica:

TIEMPO DE ESPERA (MIN)									
15	13	19	23	22	5	15	12	28	20
28	2	8	17	24	7	20	26	13	9

a) Disponga los datos en un arreglo por orden ascendente. ¿Qué comentario puede hacer acerca del tiempo de espera de los pacientes basándose para ello en su arreglo de datos?
b) Ahora construya una distribución de frecuencia utilizando 6 clases. ¿Qué otra interpretación puede dar a los datos a partir de la distribución de frecuencia?
c) Por medio de una ojiva y basándose en los datos anteriores, diga cuánto tiempo deberán esperar los pacientes en la sala de urgencias antes de recibir atención médica.

2-50 ¿Qué valor adicional tiene una distribución de frecuencia relativa, una vez construida una distribución de frecuencia?

2-51 A continuación se transcriben los pesos de una población entera de jugadores de fútbol americano de la Liga Nacional.

226	198	210	233	222	175	215	191	201	175
264	204	193	244	180	185	190	216	178	190
174	183	201	238	232	257	236	222	213	207
233	205	180	267	236	186	192	245	218	193
189	180	175	184	234	234	180	252	201	187
155	175	196	172	248	198	226	185	180	175
217	190	212	198	212	228	184	219	196	212
220	213	191	170	258	192	194	180	243	230
180	135	243	180	209	202	242	259	238	227
207	218	230	224	228	188	210	205	197	169

a) Seleccione dos muestras: una muestra de los 10 primeros elementos y otra de los 10 elementos más grandes.
b) ¿Son las dos muestras igualmente representativas de la población? De no ser así, ¿cuál es más representativa y por qué?
c) ¿En qué condiciones será la muestra de los 10 elementos más grandes tan representativa como la muestra de los 10 primeros elementos?

2-52 En la población estudiada, hay 2,000 mujeres y 8,000 hombres. Si queremos seleccionar una muestra de 250 individuos en dicha población, ¿cuántos deberán ser mujeres para que la muestra sea considerada estrictamente representativa?

2-53 El U.S. Department of Labor publica varias clasificaciones de la tasa de desempleo, junto con la tasa propiamente dicha. En fecha reciente la tasa fue de 6.8%. El departamento dio a conocer las siguientes categorías educacionales:

ESCOLARIDAD	FRECUENCIA RELATIVA (% DE LOS DESEMPLEADOS)
No terminaron la enseñanza media	.35
Recibieron diploma de enseñanza media	.31
Asistieron a la universidad pero no obtuvieron el título	.16
Recibieron un título universitario	.09
Asistieron a una escuela de posgrado pero no recibieron el título	.06
Recibieron un posgrado	.03
Total	1.00

Con los datos precedentes construya un histograma de frecuencia relativa.

2-54 Con la distribución de frecuencia relativa dada en el ejercicio 2-62, construya un histograma y polígono de frecuencias relativas. Para realizar el presente ejercicio, suponga que el límite superior de la última clase es $51.00.

2-55 Con la distribución de frecuencia relativa dada en el ejercicio 2-56 para millas por día de jogging, construya una ojiva que le ayude a estimar qué proporción de los corredores promedian 3.0 o más millas por día.

2-56 Un psicólogo deportivo, al examinar el efecto que el jogging tiene en las calificaciones de los estudiantes, reunió datos de un grupo de esos corredores. Junto con algunas otras variables, registró el número promedio de millas recorridas en un día. Recopiló sus resultados en la siguiente distribución:

MILLAS POR DIA	FRECUENCIA
1.00–1.39	32
1.40–1.79	43
1.80–2.19	81
2.20–2.59	122
2.60–2.99	131
3.00–3.39	130
3.40–3.79	111
3.80–4.19	95
4.20–4.59	82
4.60–4.99	47
5.00 en adelante	53
	927

a) Construya una ojiva que le indique aproximadamente cuántas millas por día corre un corredor medio.
b) Basándose en la ojiva que ha construido en la parte (a), ¿aproximadamente qué proporción de corredores universitarios corren un mínimo de 3.0 millas por día?

2-57 Un psicólogo al examinar el éxito que los universitarios tienen en su profesión realiza entrevistas con 100 alumnos de una universidad, mitad hombres y mitad mujeres y en ellas basa su trabajo de investigación. Comente la corrección de esta encuesta.

2-58 Si los siguientes grupos de edad están incluidos en las proporciones indicadas, ¿cuántos individuos de cada grupo deben ser incluidos en una muestra de 2,500 personas para que la muestra sea representativa?

GRUPO DE EDAD	PROPORCION RELATIVA EN LA POBLACION
12–17	.13
18–23	.34
24–29.	.24
30–35	.18
36+	.11
	1.00

2-59 La State University tiene 3 campus, cada uno con su propia escuela de administración. El año pasado los profesores de esa escuela publicaron numerosos artículos en revistas especializadas de gran prestigio, y la junta de gobierno contó los artículos como una medida de la productividad de cada departamento.

NUMERO DE REVISTAS	NUMERO DE PUBLICACIONES	CAMPUS	NUMERO DE REVISTAS	NUMERO DE PUBLICACIONES	CAMPUS
9	3	Norte	14	20	Sur
12	6	Norte	10	18	Sur
3	12	Sur	3	12	Oeste
15	8	Oeste	5	6	Norte
2	9	Oeste	7	5	Norte
5	15	Sur	7	15	Oeste
1	2	Norte	6	2	Norte
15	5	Oeste	2	3	Oeste
12	3	Norte	9	1	Norte
11	4	Norte	11	8	Norte
7	9	Norte	14	10	Oeste
6	10	Oeste	8	17	Sur

a) Construya una distribución de frecuencia y una distribución de frecuencia relativa por revista.
b) Construya una distribución de frecuencia y una distribución de frecuencia relativa por rama de la universidad.
c) Construya una distribución de frecuencia y una distribución de frecuencia relativa por el número de publicaciones (usando intervalos de 3).
d) Interprete brevemente sus resultados.

2-60 Un cuestionario de actitudes ante la educación sexual en las escuelas es enviado a una muestra aleatoria de 2,000 personas; 880 cuestionarios son llenados y devueltos al investigador. Comente los datos disponibles en esos cuestionarios en función de 5 pruebas de datos.

2-61 En cada aparato que produce una empresa de aparatos eléctricos se incluye una póliza de garantía para el cliente. Además de validar la garantía y proporcionar a la compañía el nombre y domicilio del cliente, la póliza pide otra información complementaria que se emplea en los estudios de mercado. Para cada uno de los blancos numerados de la póliza, determine las características más probables de las categorías que utilizará la compañía para registrar la información. En particular, (1) ¿serán cuantitativas o cualitativas? (2) ¿continuas o discretas? (3) ¿abiertas o cerradas? Comente brevemente el razonamiento en que se fundan sus respuestas.

Nombre________	Estado civil ____③____
Domicilio________	¿Dónde se compró el aparato?
Ciudad______ Estado___	____④____
Código Postal ______	¿Por qué compró el aparato?
Edad ① Ingreso anual __②__	____⑤____

2-62 La siguiente distribución de frecuencia resultó de un estudio sobre las cantidades que gastaron por visita los clientes de un supermercado:

CANTIDAD GASTADA	FRECUENCIA RELATIVA
\$ 0–\$ 5.99	1%
6.00–\$10.99	3
11.00–\$15.99	4
16.00–\$20.99	6
21.00–\$25.99	7
26.00–\$30.99	9
31.00–\$35.99	11
36.00–\$40.99	19
41.00–\$45.99	32
46.00 en adelante	8
Total	100%

Determine las marcas de clase (puntos medios) de cada uno de los intervalos.

2-63 Las siguientes respuestas fueron dadas por dos grupos de pacientes de hospitales; a uno de los grupos se le aplicó un nuevo tratamiento y el otro fue sometido al tratamiento ordinario. Se formuló la siguiente pregunta: "¿Qué grado de malestar está usted sintiendo?"

GRUPO 1			GRUPO 2		
Ligero	Moderado	Severo	Moderado	Ligero	Severo
Ninguno	Severo	Ligero	Severo	Ninguno	Moderado
Moderado	Ligero	Ligero	Ligero	Moderado	Moderado
Ligero	Moderado	Ninguno	Moderado	Ligero	Severo
Moderado	Ligero	Ligero	Severo	Moderado	Moderado
Ninguno	Moderado	Severo	Severo	Ligero	Moderado

Sugiera una mejor manera de mostrar estos datos. Explique por qué es mejor el método que usted propone.

2-64 El gerente de producción de una fábrica de máquinas eléctricas anotó las clasificaciones definitivas del rendimiento de los trabajadores, basándose en el total de unidades producidas, en los porcentajes de piezas rechazadas y en el total de horas trabajadas. ¿Es éste un ejemplo de datos brutos? Explique su respuesta. En caso de que sea negativa, ¿cuáles serían los datos brutos en este caso?

2-65 El jefe de un extenso departamento de administración quería clasificar las especialidades de sus 67 miembros. Pidió a Peter Wilson, un candidato al doctorado, obtener información sobre las publicaciones hechas por los miembros del profesorado. Peter recopiló los siguientes datos:

ESPECIALIDAD	MIEMBROS DEL PROFESORADO QUE PUBLICAN
Contabilidad exclusivamente	1
Mercadotecnia exclusivamente	5
Estadística exclusivamente	4
Finanzas esclusivamente	2
Contabilidad y mercadotecnia	7
Contabilidad y estadística	6
Contabilidad y finanzas	3
Mercadotecnia y finanzas	8
Estadística y finanzas	9
Estadística y mercadotecnia	21
Ninguna publicación	1
	67

Construya una distribución de frecuencia relativa para los *tipos* de especialidades. (*Sugerencia*: las categorías de su distribución serán mutuamente excluyentes, pero cualquier individuo puede caer en varias de ellas.)

2-66 Una fábrica de juguetes contrató al asesor Roberto Calderón que diseñara un nuevo programa de investigación administrativa. Con objeto de estimar las cantidades que los gerentes estarían dispuestos a invertir de sus respectivos cheques, el señor Calderón investigó los ingresos secundarios de las familias de ellos. Sus datos revelaron que ninguna familia tenía un segundo ingreso mayor que \$35,000, y varias familias al parecer no contaban con ese tipo de ingresos. En un análisis preliminar Clark decidió construir una distribución de frecuencia y una distribución de frecuencia relativa para el segundo ingreso. Pretende usar intervalos de \$5,000.

a) Prepare una distribución continua cerrada que satisfaga estas exigencias.
b) Construya una distribución continua con 6 categorías que cubra estas exigencias y que sea abierta en ambos extremos. Puede transigir en el requisito de intervalos de \$5,000 para las categorías abiertas.

2-67 La Kawahondi Computer Company recopiló datos referentes al número de entrevistas que necesitaban sus 40 vendedores para realizar una venta. A continuación se dan una distribución de frecuencia y una distribución de frecuencia relativa del número de entrevistas que se necesitan por vendedor para lograr una venta. Anote los datos faltantes:

NUMERO DE ENTREVISTAS (CLASES)	FRECUENCIA	FRECUENCIA RELATIVA
0-10	?	.05
11-20	0	?
21-30	2	?
31-40	?	?
41-50	?	.15
51-60	?	.20
61-70	5	?
71-80	?	.00
81-90	5	?
91-100	?	.00
Total	?	?

2-9 AUTOEVALUACION

Las respuestas vienen al final del libro.

V F 1. En comparación con un arreglo de datos, la distribución de frecuencia tiene la ventaja de representarlos en una forma compacta.

V F 2. Una ojiva "más que" tiene forma de S y pendiente hacia abajo y hacia la derecha.

V F 3. El histograma es una serie de rectángulos, cada uno de ancho proporcional al número de elementos que caen dentro de una clase particular de datos.

V F 4. Una observación individual se llama punto (observación) de datos, en tanto que un grupo de datos se llama tabular.

V F 5. Las clases en cualquier distribución de frecuencia relativa son exhaustivas y mutuamente excluyentes.

V F 6. Cuando una muestra contiene las características relevantes de cierta población en la misma proporción en que figuran en esta última, se dice que es una muestra representativa.

V F 7. La población es una colección de todos los elementos que estamos estudiando.

V F 8. Si quisiéramos unir los puntos medios de barras consecutivas en un histograma de frecuencias con una serie de líneas, estaríamos graficando un polígono de frecuencias.

V F 9. La información se llama datos preprocesados antes que se arregle y organice, aplicando para ello métodos estadísticos.

V F 10. Una desventaja del arreglo de datos es que no nos permite encontrar fácilmente los valores más altos y bajos en el conjunto de datos.

V F 11. Los datos discretos sólo pueden ser expresados en números enteros.

V F 12. Por lo regular, los estadísticos consideran que una distribución de frecuencia es incompleta si tiene menos de 20 clases.

V F 13. Siempre es posible construir un histograma con un polígono de frecuencias.

V F 14. La escala vertical de una ojiva para una distribución de frecuencia relativa, marca la fracción del número total de observaciones que caen dentro de cada clase.

V F 15. Un arreglo de datos se forma disponiendo los datos brutos por orden del tiempo de la observación.

V F 16. Una ojiva "menor que" tiene forma de S y pendiente hacia abajo y hacia la derecha.

V F 17. Una ventaja del histograma frente a un polígono de frecuencias consiste en que muestra con mayor claridad cada clase individual en la distribución.

18. ¿Cuál de los siguientes enunciados representa el esquema más exacto para clasificar datos?
a) Métodos cuantitativos
b) Métodos cualitativos
c) Una combinación de métodos cuantitativos y cualitativos
d) Un esquema puede determinarse sólo mediante información concreta sobre la situación.

19. ¿Cuál de los siguientes casos NO es un ejemplo de datos condensados?
a) Distribución de frecuencia c) Histograma
b) Arreglo de datos d) Ojiva

20. ¿Cuál de los siguientes enunciados sobre los rectángulos del histograma es correcto?
a) Los rectángulos son de altura proporcional al número de elementos que caen dentro de las clases.
b) Generalmente hay 5 rectángulos en cada histograma.
c) El área de un rectángulo depende del número de elementos de la clase en comparación con el número de los que hay en el resto de las clases.
d) Todos éstos.
e) a y c pero no b.

21. ¿Por qué es verdad que las clases en las distribuciones de frecuencia son exhaustivas?
a) Ninguna observación de datos caen dentro de más de una clase.
b) Siempre hay más clases que observaciones de datos.
c) Todos los datos encajan en una u otra clase.
d) Todos los anteriores.
e) a y c pero no b.

22. Al construir una distribución de frecuencia el primer paso consiste en:
a) Dividir los datos en 5 clases por lo menos.
b) Clasificar las observaciones de datos en clases y contar el número de observaciones en cada clase.
c) Escoger el tipo y número de clases para dividir los datos.
d) Ninguno de los anteriores.

23. A medida que crecen los números de observaciones y clases, la forma de un polígono de frecuencias:
a) Tiende a volverse cada vez más suave.
b) Tiende a saturarse.
c) Permanece inalterada.
d) Varía sólo si los datos se tornan más confiables.

24. ¿Cuál de los siguientes enunciados se aplica a las ojivas de frecuencia acumulativa para determinado conjunto de datos?
a) Tanto las curvas "mayores que" como las "menores que" poseen la misma pendiente.
b) Las curvas "mayores que" tienen pendiente hacia arriba y hacia la derecha.
c) Las curvas "menores que" tienen pendiente hacia abajo y hacia la derecha.
d) Las curvas "menores que" tienen pendiente hacia arriba y hacia la derecha.

25. En una ojiva construida para un conjunto particular de datos:
a) Los datos originales pueden reconstruirse siempre exactamente.
b) Siempre es posible aproximar los datos.

c) Los datos originales nunca pueden ser aproximados ni reconstruidos, pero podemos sacar conclusiones válidas respecto a ellos.
d) Ninguno de los anteriores.
e) a y b pero no c.

26. Cuando se construye una distribución de frecuencia, el número de clases que se usan depende de:
a) Número de puntos de datos.
b) Intervalo de los datos reunidos.
c) Tamaño de la población.
d) Todos los anteriores.
e) a y b pero no c.

27. ¿Cuál de los siguientes enunciados es verdadero?
a) El tamaño de una muestra nunca puede ser igual al de la población de donde se extrajo.
b) Las clases describen sólo una característica de los datos que van a ser organizados.
c) Por lo regular los estadísticos usan entre 6 y 15 clases.
d) Todos los anteriores
e) b y c pero no a.

28. Indique cuál de los siguientes números tienden a emplear los estadísticos al arreglar los datos:
a) menos de 5.
b) Entre 1 y 5.
c) Más de 30.
d) Entre 20 y 25.
e) Ninguno de los anteriores.

29. Indique cuál de las siguientes pruebas NO es una prueba útil de los datos:
a) Fuente
b) Contradicción de otra evidencia
c) Evidencia incompleta
d) Número de observaciones
e) Ninguna de las anteriores

30. Una ________________ es una colección de todos los elementos de un grupo. Una colección de algunos de esos elementos es una ________________.

31. Al dividir las observaciones de datos en clases semejantes y al contar el número de observaciones en cada clase, obtenemos una ________________.

32. Si los datos pueden tener sólo un número limitado de valores, las clases de esos datos reciben el nombre de ________________. De lo contrario, las clases se llaman ________________.

33. Una distribución de frecuencia relativa presenta las frecuencias en términos de ________________ o de ________________.

34. Una gráfica de distribuciones de frecuencia acumulativa recibe el nombre de ________________.

35. Si una colección de datos se denomina conjunto de datos, una sola observación podría llamarse ________________.

2-10 CASO CONCEPTUAL (Northern White Metals Company)

Fue a principios del otoño de 1980 cuando Dick Lennox empezó la búsqueda de trabajo que le llevaría a la Northern White Metals Company. Dick había iniciado su carrera en administración como representante de ventas de una gran empresa industrial diversificada. A fines de 1974, tras varios años de mucho éxito en ventas, abandonó una gran empresa porque no resistió la magnífica oferta que le hizo uno de sus mejores clientes, una empresa metalúrgica familiar, de tamaño mediano y muy próspera. Aceptó el puesto de director de mercadotecnia, llevando consigo algunas ideas innovadoras y el deseo vehemente de triunfar. Al final de su primer año en la empresa, ésta alcanzó un incremento récord en ventas. Dick trabajó largas horas, viajando mucho por el noreste y la región industrial del oeste medio. Las ventas siguieron creciendo, y Dick continuó adquiriendo experiencia gerencial.

Un día de verano de 1980, el presidente lo llamó a su oficina y le anunció con orgullo que sus dos hijos iban a incorporarse a la compañía. Uno sería supervisor de la planta en el departamento de estampado. El otro, que hacía poco se había graduado en una escuela de administración, que gozaba de gran prestigio en el este, sería ayudante del director de mercadotecnia y trabajaría en estrecha colaboración con Dick. De inmediato captó Dick las consecuencias que esto traería a la larga en su carrera; esa noche empezó a considerar la posibilidad de buscar otro trabajo.

Al cabo de varios meses y muchas entrevistas, Dick empezó a desanimarse. Recibió varias ofertas, algunas muy atractivas, pero ninguna correspondía a sus deseos. Y entonces, en un encuentro casual con un viejo amigo y compañero de ventas, se enteró de una pequeña firma de Nueva Inglaterra, la Northern White Metals Company, estaba buscando a una persona calificada para ocupar el puesto de director general.

Northern White Metals había iniciado operaciones a fines de la década de 1940 con el nombre de New England Metals Supply, un distribuidor de metales no ferrosos, sobre todo aluminio, cobre y latón. Los vendía en varias formas, como láminas, varilla, cable y tubería, para diversas aplicaciones industriales. Su crecimiento nunca fue impresionante, pero la compañía prosperó en el periodo de la posguerra, y las ventas y utilidades aumentaron en forma constante. En 1952 se le añadió la capacidad de producción, al comprar una vieja y abandonada fábrica textil y una prensa de extrusión de tamaño mediano, con 1,850 toneladas de aluminio. En un lapso de tres años, se vendió el negocio de cobre y latón. La Northern White Metals Company se dedicó exclusivamente a la manufactura y fabricación de los productos de aluminio extruidos, destinados principalmente a la industria de la construcción.

El proceso industrial de extrusión del aluminio comienza con la materia prima, el lingote de aluminio. Se trata esencialmente de un lingote de forma cilíndrica, la que se necesita en la prensa de extrusión. El lingote se ablanda cuando un transportador lo hace pasar por hornos a elevadas temperaturas. Luego se introduce en la prensa y se empuja con fuerte presión por un pesado troquel de acero. El resultado de ello son largas secciones de aluminio que tienen la forma deseada de sección transversal. Las secciones se extienden después para quitar las torceduras y dobleces del metal, se cortan en los tamaños deseados, se endurecen en hornos de temple, se empacan y se embarcan.

La Northern White Metals Company pronto conquistó una buena reputación por su trabajo de calidad y su entrega puntual. Para mejorar su servicio a la base de clientes, creó un departamento de anodizado, donde al metal podía dársele un acabado durable; un departamento de producción con capacidades especiales de corte, fresado, torsión y montaje; una planta pequeña de maquinaria, que ofrecía herramientas y troquel, lo mismo que capacidades especiales de reparación.

El auge de la década de 1950 y el alto potencial de crecimiento del aluminio favorecieron la proliferación de empresas industriales de tamaño mediano y pequeño. Pero pocas construyeron instalaciones primarias de manufactura, pues la mayor inversión de capital dificultaba la creación de empresas de riesgo compartido en esta área. La rápida expansión de las empresas industriales vino a intensificar la competencia. A principios de la década de 1960, una recesión y una terrible baja en los precios dejaron a la industria en una situación tal que muchas firmas nunca consiguieron recuperarse. Durante este periodo, la Northern White Metals Company operaba su departamento de producción sin obtener utilidades, siendo su finalidad principal satisfacer las necesidades de los mejores clientes. En cambio, el departamento de extrusiones prosperó y sirvió como fuente de suministro para muchas de las firmas industriales más pequeñas que se vieron obligadas a reducir radicalmente los precios para conservar su volumen de operaciones y aprovechar su capacidad instalada.

La compañía salió bien librada y, con el auge que se operó en la industria durante 1965, tuvo entonces su mejor año. Fueron ampliados los departamentos internos y la Northern White Metals Company floreció, conquistando un grupo más pequeño pero muy fiel de clientes. La depresión de 1970 hizo que incluso este grupo se contrajera, al disminuir el ritmo de la construcción comercial. Pese a ello la compañía se las arregló para seguir produciendo utilidades y hasta creció un poco, pues se buscaban nuevas aplicaciones del aluminio extruido en las empresas de alta tecnología que surgían rápidamente.

La recesión de 1974 ejerció un impacto mucho más profundo en la compañía, ya que se elevaron considerablemente los precios de los energéticos. Dado que su proceso de manufactura consumía mucha energía, los costos de Northern White Metals Company se elevaron drásticamente. La presión en las utilidades fue fuerte. Y ello, aunado a la disminución de las ventas que acompañó a la depresión económica, hizo que ese año la compañía no tuviera ganancias por primera vez desde su fundación.

Aunque la situación general de la industria comenzó a mejorar en 1975, la compañía al parecer nunca recuperó su posición anterior y tuvo una gris actuación en los siguientes años. El presidente y el principal propietario empezaron a perder interés en ella, decidiendo ponerse en contacto con un intermediario para venderla.

Encontraron a un candidato, y las negociaciones se efectuaron sin tardanza. La Northern White Metals Company fue adquirida con un intercambio de acciones comunes por la Segue Incorporated, un conglomerado diversificado de productos arquitectónicos que desde hacía mucho había querido desarrollar la capacidad de extrusión.

Fue entonces cuando Dick Lennox oyó hablar de la empresa. Concertó una cita con el director de ella, entablándose de inmediato las discusiones preliminares. Hizo un recorrido por la planta y oficinas; aunque ni una ni otras le parecieron particularmente interesantes, Dick pensó que la empresa ofrecía grandes posibilidades. Le entusiasmó profundamente la perspectiva de asumir el papel de director general.

Esa noche Dick cenó con el presidente ejecutivo de Segue. Este tenía un carácter áspero, pero relataba la historia y filosofía de su compañía con un tono casi reverente.

De repente hizo una pausa. Observó a Dick con ojos claros y penetrantes. Luego volvió a hablar.

"Lennox", le dijo con calculada deliberación, "quiero que me digas por qué debo contratarte a ti."

Dick se inclinó un poco hacia adelante y replicó sin vacilar.

"Por dos razones, señor. Yo puedo aumentar las ventas y también puedo elevar las utilidades."

Sorprendido ante una respuesta tan sencilla, el presidente ejecutivo frunció el ceño.

"Eso lo veremos, jovencito", rio entre dientes. Y luego dijo bruscamente: "Preséntate en la compañía en un plazo de tres semanas, y prepárate para empezar a trabajar."

El actual presidente de la compañía debía permanecer un año más en ella y dejar gradualmente toda la responsabilidad a Dick, el sucesor nombrado por Segue. Dirigir la compañía tal como se hallaba en esos momentos no parecía una tarea extraordinariamente difícil; en cambio, mejorar las ventas y el margen de utilidad plantearía todo un reto. Y eso era lo que Dick debía lograr.

Al cabo de tres semanas de reuniones con la fuerza de ventas, de analizar el proceso de producción y de familiarizarse con el personal de la oficina, Dick empezaba a sentirse más cómodo en su nuevo puesto. También empezó a darse cuenta del formidable reto que éste significaba. La compañía se encontraba desorganizada, los arreglos con los clientes eran informales y estaban sujetos a cambios frecuentes; los sistemas de registros también estaban muy desorganizados. Dick estaba a punto de ocuparse de todo esto cuando, el viernes por la tarde, recibió una llamada telefónica de la oficina central.

"Lennox", le espetó el presidente ejecutivo, "quiero que vengas a Nueva York el próximo lunes a hacer una presentación en la reunión de los presidentes de división. Necesitamos un análisis exhaustivo de las tendencias de los costos de producción y ventas en nuestra compañía durante los últimos años. También queremos oír algunas recomendaciones generales sobre los planes de mercadotecnia y producción para el próximo año. Danos una idea acerca de las áreas problema y las que pueden plantear algún problema. Ya sabes, la exposición que se acostumbra hacer en estas reuniones."

Dick empezó a sudar al colgar la bocina. En su antiguo empleo le bastaba pedir a su ayudante introducir esa información en la computadora, extraer algunas proyecciones y con ello quedaba listo el informe. Ahora se encontraba frente a un vetusto archivero con facturas, órdenes, registros de ventas y de producción y con un montón de todo género de información.

Dick deberá revisar un enorme acervo de datos, seleccionar la información pertinente y organizarla en una forma lógica y presentable. ¿Cómo ha de hacer todo eso?

2-11 EJERCICIO CON LA BASE DE DATOS

Cold River Company fue fundada en 1890 por la familia Mulford para satisfacer una demanda de trineos en las florecientes comunidades de veraneo de las Montañas Rocosas. Los primeros trineos eran hechos a mano por los hermanos Mulford y se vendían en ferreterías de las localidades veraniegas de Colorado.

En la década de 1920, las ventas de los trineos Rough Ridess crecieron en el noreste a medida que los turistas reconocieron la calidad del producto. Zelda y Scott Fitzgerald se deslizaron en uno de esos trineos en los Alpes suizos, y el trineo fue importado entonces como un juguete para adultos.

La situación se deterioró durante la Depresión. Las compras de juguetes y de artículos deportivos estaban relacionados con los excedentes de ingresos, por lo cual las ventas mermaron a medida que la economía decayó. Por fortuna Cold River consiguió diversificarse y empezó a fabricar duelas de barril de roble desecado que normalmente se empleaban en los trineos y a integrarse verticalmente con el whiskey ilegal que se destilaba con el agua de manantiales de las montañas rocosas. Ningún empleado fue suspendido, y la empresa siguió prosperando.

Cuando la prohibición fue levantada, Cold River volvió a producir trineos y aumentó la distribución en el oeste medio. Las ventas empezaban a incrementarse cuando estalló la Segunda Guerra Mundial. Los hermanos Mulford recurrieron otra vez a su estrategia oportunista y se dedicaron a producir cajas de armas de fuego y embalajes mientras duró la guerra.

Las ventas habían aumentado constantemente en las décadas de 1960 y 1970, a medida que el trineo Rough Rider se extendía hacia la costa oriental. En 1985, Cold River ya contaba con 50 distribuidores, y su principal línea de productos era un artículo doméstico que gozaba de un increíble reconocimiento de marca.

Los hermanos Mulford empezaron a diversificarse en la década de 1950, al fabricar pequeños juguetes de madera y varios tamaños más de trineos para entrar en el mercado infantil de gran auge. La distribución de la línea de juguetes se limitaba primordialmente a Colorado, y con frecuencia el volumen de ventas se estancaba en el punto de equilibrio.

En 1984, el conglomerado Whizbang compró una participación controladora a los nietos de los hermanos Mulford, nombrando a Joe Walsh para que depurara la empresa, aumentara las ventas y diversificase la línea de productos. El conglomerado estaba convencido de que la industria de juguetes y de tiempo libre ofrecería buenas perspectivas de crecimiento en los próximos diez años. Gracias al prestigio de Cold River, al hecho de no tener deuda alguna y a su experiencia administrativa, se pensaba ampliar las ventas e incrementar el ingreso neto a un ritmo acelerado.

Las proyecciones del U.S. Census señalaban un incremento en la natalidad para mediados de la década de 1980. Se estimaba que en 1986 habría unos 4.4 millones de nacimientos, cifra muy superior a los 3 millones por año durante la década de 1970. No sólo nacerían más niños, sino que se suponía que habría más primogénitos después de la tendencia, en esa década, a posponer la procreación. Se conjeturaba que habría un mayor ingreso disponible por niño, pues por lo regular ambos progenitores trabajaban; se preveía que tres adultos (los padres y un pariente) harían compras para cada niño. Los mensajes publicitarios se dirigirían principalmente a los que adquirían cosas para preescolares y a las posibilidades a largo plazo de los compradores de objetos para niños menores de 15 años.

Pese a esa predicción de aumento en las ventas de juguetes durante los siguientes años, Whizbang sabía que esta industria era muy arriesgada. Dos terceras partes de las compras se realizaban en las ocho semanas que precedían a la Navidad. De manera análoga, dos terceras partes de las compras al detalle hechas a los fabricantes se llevaban a cabo de agosto a octubre. Un 50% de los productos que se fabrican al año son productos nuevos; 80% de ellos no tienen éxito. La mayor parte de las compañías que fabrican juguetes se sobreponen a la merma de utilidades, basándose en los clientes estables para lograr flujo de efectivo. Dado que la mayor parte de las ventas anuales se efectúan en un periodo tan corto, los fabricantes no tienen tiempo para producir grandes cantidades de artículos de venta rápida, después que determinan cuáles productos se agotarán pronto. Las cantidades iniciales de producción de una nueva línea plantean una decisión difícil al personal de mercadotecnia, y los castigos de inventarios constituyen un fenómeno común en la industria.

Joe Walsh asumió la dirección de Cold River en este ambiente de la industria. Planeaba utilizar la empresa y sus dos extensiones de línea para subsidiar la introducción de productos nuevos durante los siguientes años. Whizbang le había dado un contrato de 5 años para incrementar las ventas en 50% y diversificar la línea de productos. Al cabo de dos años, Joe había modernizado la producción, reducido los costos e introducido un aluminio que se vendía muy bien. Estaba listo para crear una nueva línea de juguetes moldeados de plástico cuando empezaron a disminuir las ventas del trineo Rough Rider. Joe es un administrador competente y una persona muy reflexiva. Por lo regular toma buenas decisiones pero quiere que su personal le dé razones convincentes para cambiar su actitud conservadora. Hace poco contrató a Laurel McRae y Frank Grove para satisfacer la necesidad de análisis de datos y de planeación estratégica. A su juicio, una compañía de juguetes de modelo a escala debe basarse en un análisis muy refinado de mercadotecnia para tener una ventaja en la industria, y Laurel y Frank acababan de graduarse en una escuela del oeste medio que tenía gran prestigio en estas disciplinas.

En su primer día en Cold River, Frank y Laurel se saludaron con un fuerte apretón de manos mientras eran escoltados a su despacho compartido. Se habían conocido durante las entrevistas y pensaban que podían trabajar juntos. Aunque Laurel tenía sus dudas respecto a la pericia estadística de Frank, estaba segura de que sería un enlace muy útil entre el departamento de mercadotecnia y el de análisis estadístico. Frank consideraba a Laurel un poco impetuosa y falta de tacto, pero creía que podría aprovechar sus conocimientos de computación para ventaja mutua.

Fred Walker, el vicepresidente de ventas, se les unió en su prolongada pausa del café mientras se preguntaban cuándo recibirían una asignación de trabajo. Fred era el director titular de mercadotecnia en Cold River y había estado con la familia Mulford desde que empezó a vender trineos a los 18 años de edad. Había vendido en todos los territorios de la compañía y había sido gerente regional en el noreste y en el noroeste antes de ser nombrado vicepresidente de ventas en 1974. Fred estaba convencido de que la experiencia es el mejor maestro y que la educación es sólo una caparazón, algo que le quita el sabor a lo que está dentro. Desde el principio se había opuesto decididamente a la venta de la compañía. Cuando finalmente James Mulford aceptó la oferta, Fred con renuencia accedió a continuar dirigiendo las actividades de mercadotecnia de Cold River.

Fred saludó con entusiasmo a sus dos nuevos colegas y los presentó en dos despachos contiguos como "mercadólogos". Luego de aconsejarles "colgar sus diplomas para que la gente supiera que acaban de terminar sus estudios", les recordó que les había concertado una cita con Joe Walsh a las diez de la mañana.

Cuando Laurel y Frank se sentaron en el despacho del presidente ejecutivo, Joe empezó de inmediato a darles un panorama de los problemas.

"Esta es una compañía con ideas viejas. He logrado cambiar la forma en que conciben la producción, pero tardé dos años en ganarme la aceptación de la vieja guardia. Mi siguiente prioridad era la mercadotecnia, pero esta disminución de las ventas ha venido a precipitar una crisis. Los necesito a ustedes dos para examinar las órdenes de los distribuidores y averiguar que está pasando con ellos."

"De acuerdo con Fred se requiere una reunión de ventas, para sentir el pulso de mercado. Dice que los vendedores saben a que se debe la baja en las ventas, y que no necesitamos una computadora para averiguarlo."

"Sé que esta es una presentación abrupta", prosiguió Joe, "pero les ruego que examinen esos pedidos desde los tres últimos años y me den su opinión en un par de días."

Dicho esto, Joe puso sobre la mesa una cartulina de gran tamaño con pedidos. Dentro había cientos de requisiciones de órdenes, unidas con una liga en paquetes por año.

"¡Dios mío!", susurró Frank mientras pasaban por el vestíbulo cargando su pesado fargo de proyectos. "Confío que aquí tengan un perforista."

"¿Qué cosa es un perforista?", preguntó la secretaria de Frank.

PROBLEMAS Y PREGUNTAS

1. Usando los datos referentes al número de unidades vendidas por orden en los años 1983, 1984 y 1985, construya un histograma para graficar la distribución de frecuencias de cada año. Use intervalos amplios de 100 y haga que el primero quede comprendido entre 0 y 100.
2. Calcule las distribuciones de frecuencia relativa mediante los histogramas construidos en la pregunta 1.
3. ¿Qué patrón cambiante advierte en los datos año con año? ¿Cuáles son las posibles causas de ello?

UNIDADES VENDIDAS POR ORDEN, 1983

905	887	902	846	1119	1401	1299	858	702	441
874	896	923	1174	757	789	1234	927	999	1143
1004	1135	1005	1090	1270	999	812	1139	969	1348
802	656	658	482	709	426	825	362	799	731
652	345	321	396	662	365	651	321	846	580
475	543	418	390	607	562	831	598	828	655
517	343	528	722	632	655	547	848	452	333
283	285	326	176	235	198	196	190	244	208
246	258	317	270	392	242	294	336	251	214
203	200	236	238	293	294	187	256	260	246
238	339	317	293	276	250	283	293	295	202
291	303	331	257	219	285	270	338	219	367
301	274	327	201	214	208	190	329	249	230
273	340	335	214	241	257	215	262	245	320
219	274	296	297	375	247	310	228	227	248
282	226	251	204	289	284	236	232	362	312
291	314	280	230	213	256	226	251	305	333
275	219	231	174	200	264	224	261	330	185
266	191	224	225	273	276	300	229	333	196
234	288	234	214	240	190	297	202	260	307

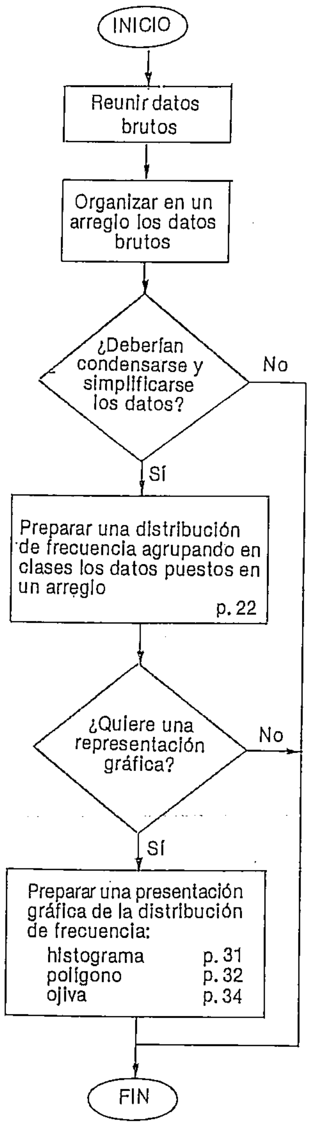
UNIDADES VENDIDAS POR ORDEN, 1984

1193	895	865	687	1337	588	1166	1136	1057	872
1326	727	1253	869	712	893	1382	1080	902	1156
990	998	919	818	770	1083	514	1047	1043	1096
1233	1129	628	1028	1159	992	935	319	659	559
484	763	557	593	633	465	329	319	871	371
380	560	411	500	454	319	649	338	525	454
524	555	507	743	524	504	876	531	403	881
839	626	421	522	721	498	330	658	192	209
290	156	271	217	203	253	202	231	303	221
178	181	219	162	235	258	228	197	228	155
201	173	249	183	309	203	252	241	272	162
184	165	220	257	142	215	256	286	188	235
242	203	192	198	212	267	181	147	269	149
313	228	290	236	228	223	258	178	273	187
245	235	206	209	171	166	205	191	198	256
262	258	249	162	300	185	146	197	158	165
268	292	239	272	198	219	191	203	196	146
218	180	228	201	156	160	229	242		

UNIDADES VENDIDAS POR ORDEN, 1985

1041	983	822	549	963	1183	982	1108	856	1025
789	844	917	565	810	1013	1261	764	958	899
704	1112	1168	780	1002	923	999	920	776	869
694	755	979	826	939	855	1060	527	869	1037
1093	1003	687	701	873	428	427	867	816	478
522	553	627	838	314	841	372	416	439	690
651	761	717	626	326	460	404	810	828	424
547	617	829	365	488	768	492	332	770	613
653	247	177	210	171	145	286	200	213	234
228	181	198	179	82	203	142	255	133	280
161	181	220	161	205	295	201	227	188	291
240	177	280	225	214	185	189	88	218	191
179	198	261	245	177	163	216	156	174	190
244	181	171	218	228	126	155	160	175	126
224	185	223	260	113	194	135	202	151	213
223									

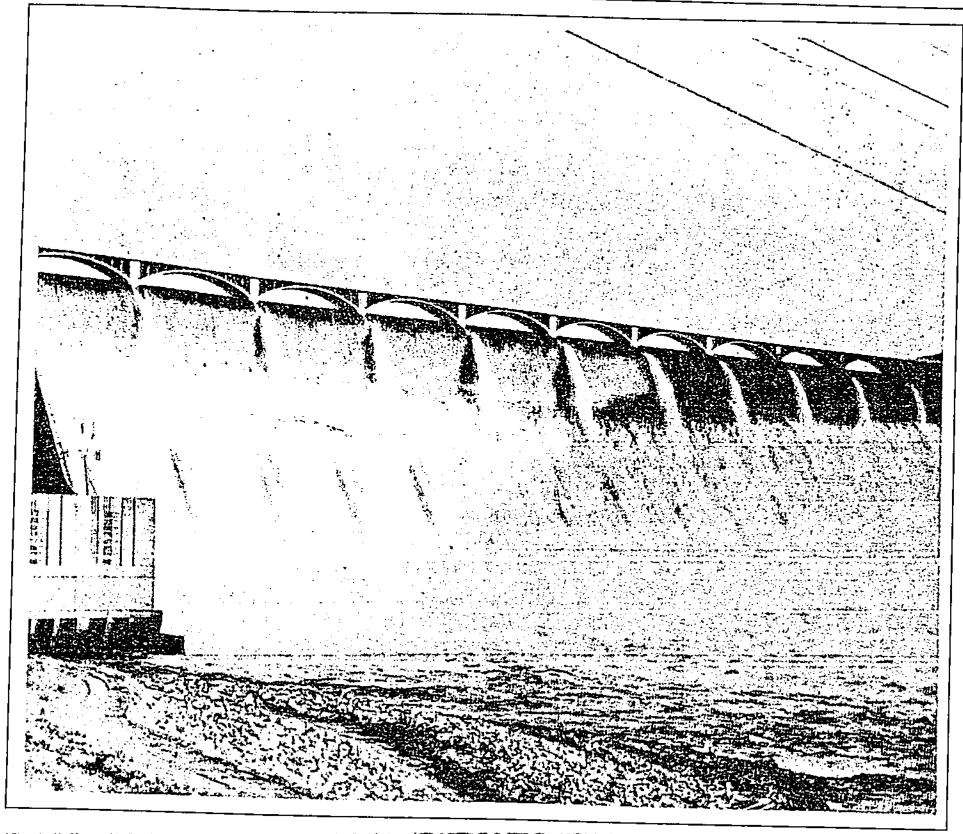
2-12 DIAGRAMA DE FLUJO



3 Medidas resumidas de las distribuciones de frecuencia

OBJETIVOS: El capítulo 3 se centra en los medios especiales con que se describe un grupo de elementos, particularmente en la forma en que las observaciones tienden a agruparse o juntarse. En este capítulo nos encontraremos con algunos términos conocidos, entre ellos el concepto de promedio. Si el entrenador del equipo de baloncesto dice que la altura promedio de los integrantes de su equipo es de 2.8 m, lo que realmente está diciendo es que la altura de los jugadores tiende a agruparse alrededor de esa cifra. Si un equipo de baloncesto tiene esa estatura, sabemos intuitivamente que hay muchas probabilidades de que tenga una temporada ganadora, y esto aun antes de estudiar formalmente la estadística. En el capítulo 3 también estudiaremos la media, la mediana y la moda —formas de medir y situar los datos.

El gerente de una planta hidroeléctrica cuenta con 10 generadores en su sistema. Necesita alguna medida del tiempo que éstos están fuera de servicio. Esa información le permitirá planear las necesidades de personal, programar el mantenimiento y organizar el servicio de respaldo. La tabla anexa contiene los datos de cada generador en el último año.



Generador	1	2	3	4	5	6	7	8	9	10
Días fuera de servicio	7	23	4	8	2	12	6	13	9	4

Al gerente le gustaría tener alguna medida de los días en que todos los generadores quedan fuera de servicio para utilizarla en su planeación. En el presente capítulo se explican algunas medidas útiles para él y para otras personas que deben hacer planes semejantes.

3-1 MAS ALLA DE TABLAS Y GRAFICAS: medidas descriptivas de las distribuciones de frecuencia

Los estadísticos resumidos describen las características de un conjunto de datos

En el capítulo 2 aprendimos a construir tablas y gráficas donde se usaban datos brutos. Las "representaciones" resultantes de las distribuciones de frecuencia nos permitieron discernir las tendencias y patrones de los datos. ¿Pero qué sucedería si necesitáramos medidas más exactas de un conjunto de datos? En ese caso, podríamos servirnos de números individuales, llamados *estadísticos resumidos*, para describir ciertas características de dicho conjunto. A partir de ellos lograremos una comprensión más precisa de los datos de la que podríamos conseguir con nuestras tablas y gráficas. Y estos números nos permitirán tomar decisiones más rápidas y satisfactorias, ya que no tendremos necesidad de consultar nuestras observaciones iniciales.

Cuatro de esas características son de suma importancia:

Mitad de un conjunto de datos

1. Medidas de tendencia central. Igual que los promedios, las medidas de tendencia central nos indican el punto medio o típico de datos que cabe esperar. También reciben el nombre de *medidas de localización*. En la figura 3-1, la localización central de la curva B se encuentra a la derecha de las localizaciones de las curvas A y C. Nótese que la localización central de la curva A es igual a la de la curva C.

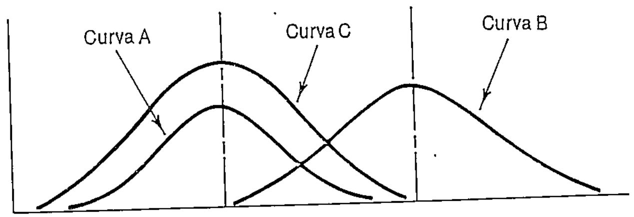


FIGURA 3-1 Comparación de la localización central de las tres curvas

Intervalo de un conjunto de datos

2. Medidas de dispersión. La *dispersión* se refiere al esparcimiento de los datos, o sea al grado de dispersión de las observaciones. En el capítulo 2, hemos estudiado una medida de dispersión denominado intervalo. Este indica la distancia que existe entre el punto de datos más bajo y el más alto. Adviértase que la curva A en la figura 3-2 tiene una dispersión mayor que la curva B.

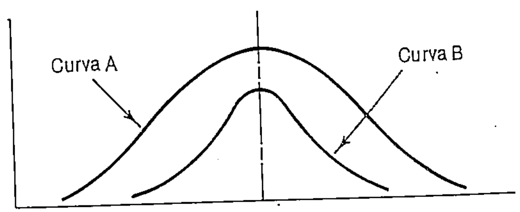


FIGURA 3-2 Comparación de la dispersión de dos curvas

Simetría de un conjunto de datos

3. **Medidas de sesgo (asimetría).** Las curvas que representan las observaciones de datos en el conjunto de datos pueden ser simétricas o asimétricas (sesgadas). Las curvas *simétricas*, como la de la figura 3-3, son tales que una línea vertical trazada desde la cumbre de la curva al eje horizontal dividirá el área de la curva en dos partes iguales. Cada parte es la imagen de espejo de la otra.

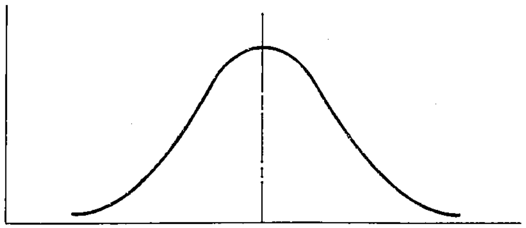


FIGURA 3-3
Curva simétrica

Sesgo (asimetría) de un conjunto de datos

Las curvas A y B en la figura 3-4 son curvas *sesgadas*. Y lo son porque los valores de sus distribuciones de frecuencia se concentran en el extremo inferior o en el extremo superior de la escala de medición situada sobre el eje horizontal. Los valores no tienen una distribución igual. La curva A está sesgada a la derecha (también se dice que tiene asimetría *positiva*), pues disminuye gradualmente hacia el extremo superior de la escala. Lo contrario sucede con la curva B. Está sesgada a la izquierda (tiene asimetría *negativa*) dado que disminuye gradualmente hacia el extremo inferior de la escala.

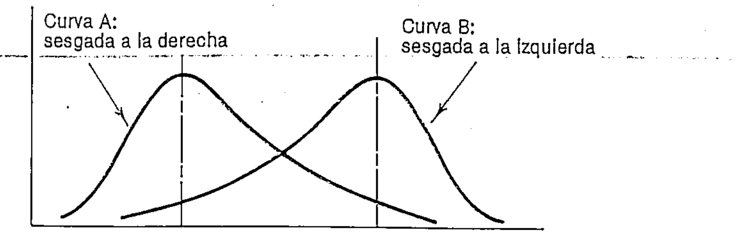


FIGURA 3-4
Comparación de dos curvas sesgadas

La curva A podría representar la distribución de frecuencia del número de existencias de que diariamente dispone una empresa mayorista de frutas. La curva estaría sesgada a la derecha, con muchos valores situados en el extremo inferior y pocos en el extremo superior, ya que podría haber una rotación rápida de inventario. De manera análoga, la curva B podría representar la frecuencia del número de días que un corredor de bienes raíces necesita para vender una casa. Sería asimétrica a la izquierda, con muchos valores concentrados en el extremo superior y pocos en el extremo inferior, puesto que el inventario de casas, muestra una rotación muy lenta.

Grado de pico de un conjunto de datos

4. **Medidas de curtosis.** Cuando medimos la *curtosis* de una distribución, estamos midiendo su grado de pico. En la figura 3-5, por ejemplo, las curvas A y B difieren tan sólo por el hecho de que una tiene un pico mayor que la otra. Ambas poseen la misma localización y dispersión, y ambas son también simétricas. Los estadísticos dicen que las dos curvas tienen diferentes grados de curtosis.

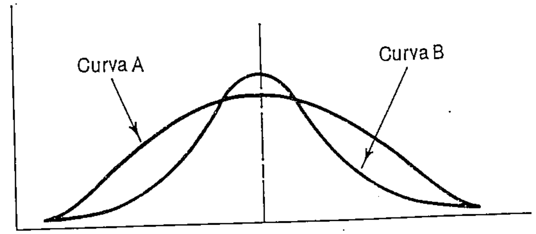


FIGURA 3-5
Dos curvas con la misma localización central pero con diferente curtosis

Existen muchos grados diferentes de curtosis, pero los estadísticos normalmente utilizan tres clases generales. Una curva como la de la figura 3-6 se llama *mesocúrtica*; una curva que tenga más pico, como la de la figura 3-7, recibe el nombre de *leptocúrtica*; y una curva que tenga menos pico, como la de la figura 3-8, se denomina *platicúrtica*.

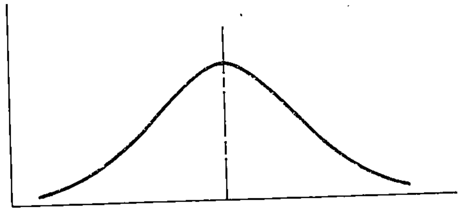


FIGURA 3-6
Curva mesocúrtica (*meso*, palabra griega que significa "mitad")

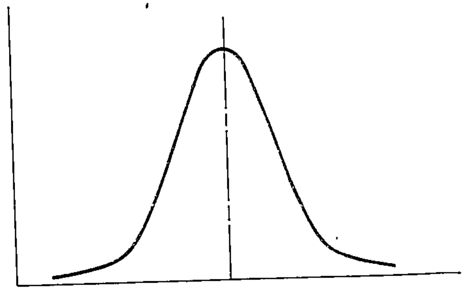


FIGURA 3-7
Curva leptocúrtica (*lepto*, palabra griega que significa "esbelto")

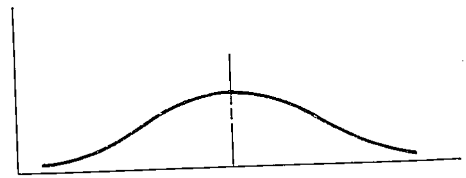


FIGURA 3-8
Curva platicúrtica (*platos*, palabra griega que significa "ancho o plano")

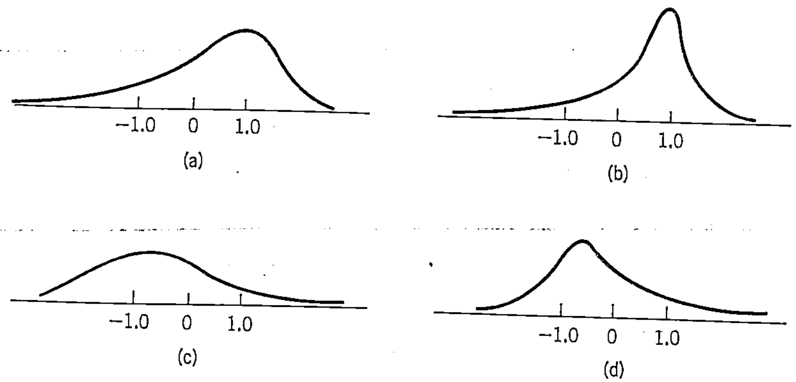
Ahora que hemos descrito brevemente estas características de las distribuciones de frecuencia, podemos describir de manera más pormenorizada tres *medidas comunes de tendencia central*: la *media*, la *mediana* y la *moda*.

EJERCICIOS

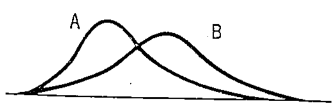
3-1 Dé ejemplos de las siguientes distribuciones, A y B:
a) A: simétrica, mesocúrtica, intervalo de −1.0 a +1.0, tendencia central de 0.0.
   B: simétrica, platicúrtica, intervalo de −1.5 a +1.5, tendencia central de 0.0.
b) A: sesgada a la izquierda, mesocúrtica, intervalo de −1.0 a +1.0, pico de +0.5.
   B: simétrica, mesocúrtica, intervalo de −1.0 a +1.0, tendencia central de 0.0.
c) A: simétrica, leptocúrtica, intervalo de −0.5 a +0.5, tendencia central de 0.0.
   B: sesgada hacia la derecha, leptocúrtica, intervalo de −1.0 a +1.0, pico de −0.5.

3-2 Dibuje 3 curvas, todas ellas simétricas y con la misma dispersión, pero con las siguientes localizaciones centrales: a) 0.0 b) 1.0 c) −1.0

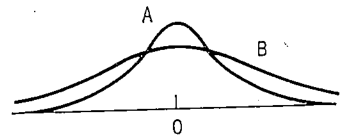
3-3 A continuación se muestran 4 curvas de distribución. Para cada una, indique su pico, su grado de curtosis, si es simétrica, si tiene sesgo positivo o negativo.



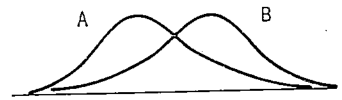
3-4 En las siguientes distribuciones indique qué distribución:
a) Tiene el mayor valor promedio.
b) Tiene mayores probabilidades de producir un valor pequeño que un valor grande.
c) Es la mejor representación de la distribución de edades en un concierto de rock.
d) Es la mejor representación de la distribución de las veces que debe usted esperar en el consultorio de un médico.



Para las dos distribuciones siguientes, indique cuál distribución:
e) Tiene valores más uniformemente distribuidos en el intervalo de los valores posibles.
f) Tiene mayores probabilidades de producir un valor cercano a 0.
g) Tiene mayores probabilidades de producir valores grandes que valores pequeños.



3-5 Si las dos curvas siguientes representan la distribución de las calificaciones obtenidas en 2 pruebas por un grupo de estudiantes, ¿cuál prueba parece ser más difícil para ellos, A o B? Explique su respuesta.



3-2 UNA PRIMERA MEDIDA DE TENDENCIA CENTRAL: la media aritmética

La mayor parte de las veces cuando nos referimos al "promedio" de algo, estamos hablando de la media aritmética. Ello sucede en tales casos como la temperatura invernal promedio de la Ciudad de Nueva York, la vida promedio de una batería de luz intermitente y la producción promedio de maíz en un acre de tierra.

TABLA 3-1 Tiempo ocioso de los generadores en la estación de Lake Ico

GENERADOR	1	2	3	4	5	6	7	8	9	10
DIAS FUERA DE SERVICIO	7	23	4	8	2	12	6	13	9	4

La media aritmética es un promedio

La tabla 3-1 repite los datos del ejemplo con que se inicia este capítulo. Esos datos representan el número de días que los generadores están fuera de servicio debido a un mantenimiento insatisfactorio o a fallas. Para calcular la media aritmética, sumamos los valores y los dividimos después entre el número de observaciones.

$$\text{Media aritmética} = \frac{7+23+4+8+2+12+6+13+9+4}{10}$$
$$= \frac{88}{10}$$
$$= 8.8 \text{ días}$$

En este periodo de un año, los generadores permanecieron fuera de servicio un promedio de 8.8 días. Conociendo esta cifra, el gerente de la planta hidroeléctrica cuenta con una medida razonable del comportamiento de *todos* sus generadores.

Símbolos convencionales

Las características de una muestra se llaman estadísticos

Con objeto de escribir ecuaciones para estas medidas de las distribuciones de frecuencia, necesitamos aprender las notaciones matemáticas que emplean los estadísticos. Una *muestra* de una población está constituida por n observaciones (una n minúscula) con una media de $\bar{x}$ (léase barra x). Recuérdese que las medidas que obtenemos para una muestra se denominan *estadísticos*.

Las características de una población se llaman parámetros

La notación es diferente cuando estamos calculando las medidas de la *población* entera; es decir, para el grupo que contiene todos los elementos que estamos describiendo. La media de una población se representa mediante μ que es la letra griega *my* (mu). El número de elementos en una población se denota con la letra cursiva mayúscula N. Generalmente, en estadística empleamos las letras redondas para representar la información referente a las muestras y las letras griegas para representar la información acerca de la población.

Cálculo de la media a partir de datos no agrupados

Obtención de la media de la población y de la muestra

En el ejemplo, el promedio de 8.8 días será μ (la media de la población) si la población de generadores es exactamente 10. Será $\bar{x}$ (la media muestral) si los 10 generadores son una muestra extraída de una población mayor. Para escribir las fórmulas de esas dos medias, combinamos nuestros símbolos matemáticos y los pasos que se requieren para determinar la media aritmética. Si sumamos los valores de las observaciones y dividimos la suma entre el número de observaciones, obtendremos:

Media de la población

Suma de valores de todas las observaciones

$$\mu = \frac{\Sigma x}{N} \qquad [3\text{-}1]$$

Número de elementos en la población

y:

Media muestral

Suma de los valores de todas las observaciones

$$\bar{x} = \frac{\Sigma x}{n} \qquad [3\text{-}2]$$

Número de elementos en la muestra

Puesto que μ es la *media aritmética de la población*, nos servimos de N para indicar que dividimos entre el número de observaciones o elementos de la población. De manera análoga, $\bar{x}$ es la *media aritmética de la muestra* y n es el número de observaciones en la muestra. La letra griega sigma, Σ, indica que todos los valores de x están sumados juntos.

He aquí otro ejemplo: la tabla 3-2 contiene el incremento percentil en las puntuaciones verbales de la prueba SAT (test de aptitudes académicas) obtenidas por siete estudiantes que asisten a un curso preparatorio de ese test.

TABLA 3-2 Incremento percentil en las puntuaciones verbales de la prueba SAT

ESTUDIANTE	1	2	3	4	5	6	7
INCREMENTO	9	7	7	6	4	4	2

Los datos se ponen en un arreglo por orden descendente. Suponemos que hay demasiados estudiantes en el curso como para someterlos a un examen. Por tanto, recurrimos a nuestra muestra y obtenemos la media del modo siguiente:

$$\bar{x} = \frac{\Sigma x}{n} \qquad [3\text{-}2]$$

$$= \frac{9+7+7+6+4+4+2}{7}$$

$$= \frac{39}{7}$$

$$= 5.6 \text{ puntos por estudiante} \leftarrow \text{media muestral}$$

Cómo trabajar con datos no agrupados

Nótese que, para calcular esta media, hemos sumado todas las observaciones por separado, en ningún orden especial. A esto los estadísticos lo llaman datos *no agrupados*. Los cálculos no fueron difíciles puesto que nuestro tamaño de la muestra fue pequeño. Pero supongamos ahora que estamos trabajando con el peso de 5,000 cabezas de ganado y preferimos no sumar por separado las observaciones (puntos) de datos. O supongamos que tenemos acceso sólo a la distribución de frecuencia de los datos, no a cada observación individual. En tales casos, necesitaremos otro método para calcular la media aritmética.

Cálculo de la media a partir de datos agrupados

Cómo trabajar con datos agrupados

Una distribución de frecuencia está compuesta de datos que se agrupan por clases. Cada valor de una observación cae en algún lugar de una de las clases. A diferencia del ejemplo del test de aptitudes académicas (SAT), no conocemos los valores individuales de cada observación. Supongamos que tenemos una distribución de frecuencia (que aparece en la tabla 3-3) sobre el promedio de los saldos mensuales de las cuentas de cheques de 600 clientes en una sucursal. Con la información contenida en esta tabla podemos calcular fácilmente una *estimación* del valor de la media

Estimación de la media

de los datos agrupados. Se trata de una estimación porque no usamos las 600 observaciones de datos de la muestra. De haber utilizado los datos originales no agrupados, habríamos obtenido el valor real de la media pero sólo después de haber promediado 600 valores individuales. Para facilitar el cálculo hemos de renunciar a la exactitud.

TABLA 3-3 Saldos mensuales promedio de 600 clientes

CLASE (EN DOLARES)	FRECUENCIA
0– 49.99	78
50.00– 99.99	123
100.00–149.99	187
150.00–199.99	82
200.00–249.99	51
250.00–299.99	47
300.00–349.99	13
350.00–399.99	9
400.00–449.99	6
450.00–499.99	4
	600

Cálculo de la media

Para encontrar la media aritmética de los datos agrupados, primero calculamos el punto medio de cada clase (la marca de clase). Si queremos que las marcas de clase vengan en centenas enteras, las redondeamos. Así, por ejemplo, la marca de la primera clase se convierte en 25.00 en vez 24.995. Después multiplicamos cada marca de clase por la frecuencia de observaciones en esa clase, sumamos todos estos resultados y dividimos el total entre el número total de observaciones de la muestra. La fórmula es ésta:

$$\bar{x} = \frac{\Sigma(f \times x)}{n} \qquad [3\text{-}3]$$

donde:

- $\bar{x}$ es la media muestral
- Σ es el símbolo que significa "la suma de"
- f es la frecuencia (número de observaciones) en cada clase
- x representa la marca de clase de cada clase de la muestra
- n es el número de observaciones en la muestra

La tabla 3-4 ilustra cómo calcular la media aritmética con nuestros datos agrupados, aplicando la ecuación 3-3.

TABLA 3-4 Cálculo de la media muestral aritmética con datos agrupados en la tabla 3-3

CLASE (EN DOLARES) (1)	MARCAS DE CLASE (2)		FRECUENCIA (3)		$f \times x$ (3) × (2)
0– 49.99	25.00	×	78	=	1,950
50.00– 99.99	75.00	×	123	=	9,225
100.00–149.99	125.00	×	187	=	23,375
150.00–199.99	175.00	×	82	=	14,350
200.00–249.99	225.00	×	51	=	11,475
250.00–299.99	275.00	×	47	=	12,925
300.00–349.99	325.00	×	13	=	4,225
350.00–399.99	375.00	×	9	=	3,375
400.00–449.99	425.00	×	6	=	2,550
450.00–499.99	475.00	×	4	=	1,900
			$\Sigma f = n =$ 600		85,350 ← $\Sigma(f \times x)$

$$\bar{x} = \frac{\Sigma(f \times x)}{n} \qquad [3\text{-}3]$$

$$= \frac{85{,}350}{600}$$

$= 142.25$ ← media muestral (en dólares)

En nuestra muestra de 600 clientes, el promedio del saldo mensual de las cuentas de cheques es de \$142.25. Esta es nuestra aproximación de la distribución de frecuencia. Nótese que, como no conocíamos todas las observaciones de los datos en la muestra, supusimos que cada valor de una clase era igual a su marca de clase. Así pues, nuestros resultados sólo pueden aproximar el saldo mensual promedio.

Comparación de la media estimada y de la media verdadera

Comparemos una media aproximada calculada con los datos agrupados, obteniendo la media real de datos no agrupados. Examinemos el ejemplo presentado en las tablas 3-5 y 3-6 que registran las nevadas anuales (en pulgadas) durante 20 años en Harlan, Kentucky. Si usamos datos no agrupados, la nevada anual promedio será 21.65 pulgadas. Si nos servimos de datos agrupados, el promedio estimado será de 21.5. La diferencia es pequeña. Y cuando el número de observaciones es grande, apreciaremos la comodidad que nos brindan los datos agrupados.

Codificación

Asignación de códigos a las marcas de clase

Cuando tenemos que realizar manualmente las operaciones aritméticas, podemos simplificar aún más nuestro cálculo de la media a partir de datos agrupados. Aplicando un método llamado *codificación*, eliminamos el problema de marcas de clases grandes o inadecuadas. En vez de usar las verdaderas marcas de clase para efectuar los cálculos, es posible asignar enteros consecutivos de valor pequeño (números enteros) denominados *códigos* a cada una de las marcas de clase. El entero cero puede ser asignado en cualquier lugar, pero para mantener pequeños los enteros asignaremos cero a la marca de clase en la *mitad* (o en el entero más cercano a la mitad) de la distribución de frecuencia. Después podemos asig-

TABLA 3-5 Nevadas anuales en Harlan, Kentucky

AÑO	NEVADA (EN PULGADAS)	AÑO	NEVADA (EN PULGADAS)
1967	23	1977	12
1968	8	1978	28
1969	14	1979	8
1970	31	1980	36
1971	5	1981	16
1972	26	1982	9
1973	11	1983	42
1974	27	1984	30
1975	32	1985	7
1976	46	1986	22
			433 ← Total de nevadas

$$\bar{x} = \frac{\Sigma x}{n} \quad [3\text{-}2]$$

$$= \frac{433}{20}$$

$$= 21.65 \leftarrow \text{Promedio anual de nevadas}$$

TABLA 3-6 Nevadas anuales en Harlan, Kentucky

CLASE (DATOS AGRUPADOS) (1)	MARCA DE (x) CLASE (2)		FRECUENCIA (f) (3)		f × x (3) × (2)
0– 7	3.5	×	2	=	7.0
8–15	11.5	×	6	=	69.0
16–23	19.5	×	3	=	58.5
24–31	27.5	×	5	=	137.5
32–39	35.5	×	2	=	71.0
40–47	43.5	×	2	=	87.0
					430.0 ← Σ(f × x)

$$\bar{x} = \frac{\Sigma(f \times x)}{n} \quad [3\text{-}3]$$

$$= \frac{430}{20}$$

$$= 21.5 \leftarrow \text{Promedio anual de nevadas}$$

nar enteros negativos a los valores más pequeños que esa marca de clase y enteros positivos a los valores más grandes como sigue:

Clase	1–5	6–10	11–15	16–20	21–25	26–30	31–35	36–40	41–45
Código (u)	−4	−3	−2	−1	0	1	2	3	4
					↑ x_0				

Cálculo de la media de datos agrupados, usando códigos

Simbólicamente, los estadísticos se sirven de x_0 para representar la marca de clase a la cual se asigna el código 0 y u para designar las marcas codificadas. La siguiente fórmula se aplica al determinar la media muestral que utiliza códigos:

$$\bar{x} = x_0 + w\frac{\Sigma(u \times f)}{n} \quad [3\text{-}4]$$

donde:

- $\bar{x}$ = media de la muestra
- x_0 = valor de la marca de clase a la cual se asigna el código 0
- w = amplitud numérica del intervalo de clase
- u = código asignado a cada clase
- f = frecuencia o número de observaciones en cada clase
- n = número total de observaciones en la muestra

No se olvide que $\Sigma(u \times f)$ significa simplemente que 1) multiplicamos u por f para cada clase en la distribución de frecuencia y 2) sumamos todos esos productos. La tabla 3-7 muestra cómo codificar las marcas de clase y obtener la media muestral. El resultado es el mismo que cuando calculamos la media de los datos agrupados sin codificación (ilustrado esto en la Tabla 3-6).

TABLA 3-7 Nevadas anuales en Harlan, Kentucky

CLASE (1)	MARCA DE CLASE(x) (2)	CODIGO(u) (3)		FRECUENCIA (f) (4)		u × f (3) × (4)
0–7	3.5	−2	×	2	=	−4
8–15	11.5	−1	×	6	=	−6
16–23	19.5 ← x_0	0	×	3	=	0
24–31	27.5	1	×	5	=	5
32–39	35.5	2	×	2	=	4
40–47	43.5	3	×	2	=	6
				Σf = n = 20		5 ← Σ(u × f)

$$\bar{x} = x_0 + w\frac{\Sigma(u \times f)}{n} \quad [3\text{-}4]$$

$$= 19.5 + (8)\left(\frac{5}{20}\right)$$

$$= 19.5 + 2$$

$$= 21.5 \leftarrow \text{Promedio anual de nevadas}$$

Ventajas y desventajas de la media aritmética

Ventajas de la media

La media aritmética, en cuanto número individual que representa un conjunto entero, ofrece importantes ventajas. Primero, su concepto lo conocen casi todos y es intuitivamente claro. Segundo, todo conjunto de datos posee una media. Es una medida calculable, y es especial porque todo conjunto de datos tiene una y sólo una media. Por último, la media sirve para realizar procedimientos estadísticos como la comparación de las medias a partir de varios conjuntos de datos (procedimiento que llevaremos a cabo en el Capítulo 9).

TABLA 3-8 Tiempos de los miembros del equipo de pista en la carrera de la milla

MIEMBRO	1	2	3	4	5	6	7
TIEMPO EN MINUTOS	4.2	4.3	4.7	4.8	5.0	5.1	9.0

Desventajas de la media

Sin embargo, por ser una medida estadística, la media aritmética tiene desventajas que es preciso conocer. **Primero**, aunque es confiable porque refleja todos los valores en el conjunto de datos, puede ser afectada por los valores extremos que no sean representativos del resto de ellos. Nótese que, si los siete miembros de un equipo de pista tienen los tiempos en la carrera de la milla que aparecen en la tabla 3-8, el tiempo promedio será:

$$\mu = \frac{\Sigma x}{N} \qquad [3\text{-}1]$$

$$= \frac{4.2 + 4.3 + 4.7 + 4.8 + 5.0 + 5.1 + 9.0}{7}$$

$$= \frac{37.1}{7}$$

$$= 5.3 \text{ minutos} \leftarrow \text{media de la población}$$

Si calculamos un tiempo promedio para los seis primeros miembros y si excluimos el valor 9.0, la respuesta será unos 4.7 minutos. El *valor extremo* de 9.0 distorsiona el valor que obtenemos de la media. Sería más representativo calcular la media *sin* incluirlo.

Un **segundo** problema de la media es el mismo que hemos encontrado en los 600 saldos de las cuentas de cheques: es tediosa obtenerla debido a que usamos todas las observaciones de datos al hacerlo (a menos que apliquemos el método rápido consistente en emplear datos agrupados para aproximar la media).

La **tercera** desventaja consiste en que somos incapaces de calcular la media para un conjunto de datos que tengan clases abiertas en el extremo superior o inferior de la escala. Supóngase que los datos de tabla 3-8 han sido dispuestos en la distribución de frecuencia que aparece en la tabla 3-9. No podríamos calcular un valor promedio de esos datos, debido a la clase abierta "5.4 y más". No hay manera de conocer si el valor es 5.4, cercano a 5.4 o muy por encima de 5.4.

TABLA 3-9 Tiempos de los miembros del equipo de pista en la carrera de la milla

CLASE EN MINUTOS	4.2–4.5	4.6–4.9	5.0–5.3	5.4 en adelante
FRECUENCIA	2	2	2	1

Necesidad de emplear los promedios con prudencia

Problemas de los promedios

Una de las anécdotas más antiguas sobre el empleo incorrecto de los promedios es el de un reclutador cuya misión consistía en seleccionar pilotos para la fuerza aérea. Debía asegurarse de que el piloto promedio midiera 1.82 m de altura. Se cuenta que salió en busca de reclutas y que encontró a dos bastante extraños: uno medía apenas 1.21 y el otro 2.43. Desde luego ninguno de ellos cumplía con los requisitos, pues el más pequeño no alcanzaba a ver por arriba del tablero de instrumentos en la cabina; el más alto ni siquiera cabía en la cabina. Pero pese a ello tenían un promedio de 1.82 metros. Hemos de ser sensibles a este tipo de cosas cada vez que usemos una medida de tendencia central para describir los datos. No obstante, en muchos casos el engaño estadístico es mucho menos evidente que en el ejemplo del reclutador de pilotos.

Otro ejemplo de una interpretación errónea de los promedios se ilustra al señalar cómo las personas, a quienes preocupa mucho el uso que se da a los impuestos, se enfadan ante la utilización aparentemente escasa de las aulas en las universidades públicas. A veces oímos afirmaciones como la siguiente: "En la universidad de mi estado la utilización promedio del aula es apenas 34%." ¿Cómo hemos de reaccionar ante tales afirmaciones? La cifra que se cita como media y que incluye las horas posibles de clase puede tener poco o nulo valor en el proceso educativo. Por ejemplo, las clases que se imparten ya muy noche (digamos después de las 9 p.m.) tal vez sean improductivas en lo tocante al aprendizaje por la fatiga. Y, por otra parte, las clases programadas antes de las 8 de la mañana acaso no sean factibles debido a la interrupción del periodo normal del sueño y a la falta de transporte adecuado en esas horas. Por último, las clases programadas los sábados y domingos pueden suscitar resistencia a causa de tradiciones religiosas. Cuando se atiende a todos los factores que acabamos de mencionar, se llega a la conclusión de que la utilización de las aulas puede ser significativamente mayor al 34%; pero mientras no conozcamos el fundamento del cálculo del promedio no podemos reaccionar de manera inteligente frente a una cifra publicada.

EJERCICIOS

3-6 Una guardería es una institución elegible para recibir un subsidio destinado a los servicios sociales del condado, a condición de que la edad promedio de sus niños no llegue a 9. Si los datos siguientes representan la edad de todos los niños que actualmente asisten a ella, ¿llena los requisitos para recibir el subsidio?

8, 5, 9, 10, 9, 12, 7, 12, 13, 7, 8

3-7 La guardería del ejemplo anterior puede continuar siendo subvencionada por la oficina de servicios sociales del condado, mientras el ingreso anual promedio de las familias cuyos hijos asisten a esa institución no lleguen a $12,500. El ingreso familiar de los padres de los niños es:

$14,500, $15,600, $12,500, $8,600, $ 7,800,
$ 6,500, $ 5,900, $10,200, $8,800, $14,300, $13,900

a) ¿Llena esta institución los requisitos para recibir apoyo financiero del condado?
b) Si la respuesta a (a) es negativa, ¿cuánto debe disminuir el ingreso familiar para cumplir con esa condición?
c) Si la respuesta a (a) es afirmativa, ¿cuánto puede aumentar el ingreso familiar promedio sin que la institución pierda su elegibilidad?

3-8 Los datos que se transcriben en seguida representan las edades de los pacientes que ingresan a un pequeño hospital el 28 de febrero de 1987:

85	75	66	43	40
88	80	56	56	67
89	83	65	53	75
87	83	52	44	48

a) Construya una distribución de frecuencia con las clases 40-49, 50-59, etc.
b) Obtenga la media muestral con la distribución de frecuencia.
c) Calcule la media muestral con los datos brutos.
d) Compare (b) y (c) y comente su respuesta.

3-9 La distribución de frecuencia que se da en seguida, representa los pesos en kilogramos de una muestra de paquetes que en el mes pasado transportó una pequeña compañía de transporte aéreo.

CLASE	FRECUENCIA	CLASE	FRECUENCIA
10.0-10.9	1	15.0-15.9	11
11.0-11.9	4	16.0-16.9	8
12.0-12.9	6	17.0-17.9	7
13.0-13.9	8	18.0-18.9	6
14.0-14.9	12	19.0-19.9	2

a) Calcule la media muestral aplicando el método de marcas de clase (Ecuación 3-3).
b) Calcule la media muestral empleando el método de codificación (Ecuación 3-4), asignado 0 a la cuarta clase.
c) Repita (b) asignando 0 a la sexta clase.
d) Explique por qué sus respuestas en (b) y (c) son iguales.

3-10 Una fábrica de muebles tiene un contrato de crédito revolvente con un banco del país. El préstamo mostró los siguientes saldos mensuales finales en el año pasado:

Ene.	$121,300	Abr.	$72,800	Jul.	$58,700	Oct.	$52,800
Feb.	$112,300	May.	$72,800	Ago.	$61,100	Nov.	$49,200
Mar.	$ 72,800	Jun.	$57,300	Sep.	$50,400	Dic.	$46,100

La compañía es elegible para recibir una tasa reducida de interés si su saldo mensual promedio es mayor de $65,000; ¿cumple con tal requisito?

3-11 Un fabricante de cosmético compró hace poco una máquina para llenar frascos con agua de colonia de 3 onzas. Para probar la precisión de fijación de volumen de la máquina, se llenaron 18 frascos a manera de ensayo. Los volúmenes resultantes (en onzas) en los ensayos fueron los siguientes:

3.02	2.89	2.92	2.84	2.90	2.97	2.95	2.94	2.93
3.01	2.97	2.95	2.90	2.94	2.96	2.99	2.99	2.97

La compañía normalmente no recalibra la máquina de llenado para esta agua de colonia, si el volumen promedio se halla entre .04 de 3.00 onzas; ¿debe ser recalibrada?

3-12 El gerente de producción de Hinton Press está determinando el tiempo promedio que se requiere para fotografiar una plancha de impresión. Usando un cronómetro y observando, los aparatos que hacen las placas obtiene los siguientes tiempos (en segundos):

20.4	20.0	22.2	23.8	21.3	25.1	21.2	22.9	28.2	24.3
22.0	24.7	25.7	24.9	22.7	24.4	24.3	23.6	23.2	21.0

Un promedio de tiempo por placa de menos de 23.0 segundos indica una productividad satisfactoria. ¿Debería preocuparse el gerente de producción?

3-13 National Tire Company tiene fondos de reserva en valores cotizados a corto plazo. El saldo diario final (en millones) de la cuenta en esos valores, correspondiente a dos semanas, se muestra a continuación:

Semana 1	$1.973	$1.970	$1.972	$1.975	$1.976
Semana 2	1.969	1.892	1.893	1.887	1.895

¿Cuál era la cantidad promedio (media) invertida en valores cotizados durante
a) la primera semana?
b) la segunda semana?
c) el periodo de dos semanas?
d) Un saldo promedio durante las dos semanas de más de $1,970 hace que la compañía pueda recibir tasas especiales de interés. ¿Cumple con el requisito?
e) si la respuesta a (d) es menor de $1,970 millones, ¿cuánto deberá aumentar la cantidad invertida el último día para que la compañía reciba las tasas especiales de interés?
f) Si la respuesta a (d) es mayor de $1,970 millones, ¿cuánto puede sacar el tesorero de los fondos de reserva en el último día sin que por ello la compañía deje de ser elegible para las tasas especiales de interés?

3-14 Miguel Tejero viaja al este de los Estados Unidos comisionado por un editor de libros de texto. Le pagan una comisión que depende del volumen de ventas. Sus ganancias trimestrales durante los 3 últimos años se transcriben a continuación.

	PRIMER TRIMESTRE	SEGUNDO TRIMESTRE	TERCER TRIMESTRE	CUARTO TRIMESTRE
Año 1	$10,000	$ 5,000	$25,000	$15,000
Año 2	20,000	10,000	20,000	10,000
Año 3	30,000	15,000	45,000	50,000

a) Calcule por separado las ganancias promedio del señor Tejero en cada uno de los cuatro trimestres.
b) Calcule por separado las ganancias trimestrales promedio en cada uno de los tres años.
c) Muestre que la media de los 4 números que encontró en la parte a) es igual a la media de los 3 números que encontró en la parte b). Más aún, muestre que esos número son iguales a la media de los 12 números de la base de datos. (Este es el ingreso trimestral promedio del señor Tejero durante 3 años.)

3-15 Liliana Durán ha sido presidenta del comité de bibliotecas del condado desde hace 10 años. Afirma que durante su gestión ha manejado el presupuesto de la biblioteca móvil mejor que su antecesor. He aquí los datos de reparaciones a la biblioteca móvil durante 15 años:

AÑO	PRESUPUESTO DE LA CD.	AÑO	PRESUPUESTO DE LA CD.	AÑO	PRESUPUESTO DE LA CD.
1984	$30,000	1979	$24,000	1974	$30,000
1983	28,000	1978	19,000	1973	20,000
1982	25,000	1977	21,000	1972	15,000
1981	27,000	1976	22,000	1971	10,000
1980	26,000	1975	24,000	1970	9,000

a) Calcule el presupuesto anual promedio de los últimos cinco años (1980-1984).
b) Calcule el presupuesto anual promedio de los primeros años de Liliana en su cargo de presidenta (1975-1979).
c) Calcule el presupuesto anual promedio de los 5 años antes de ser elegida (1970-1974).
d) Basándose en las respuestas que obtuvo en las partes (a), (b) y (c), ¿cree que haya habido una tendencia descendente o ascendente en el presupuesto anual? ¿Ha ahorrado Liliana dinero al condado?

3-3 UNA SEGUNDA MEDIDA DE TENDENCIA CENTRAL: la media ponderada

Una media ponderada

La media ponderada nos permite obtener un promedio que tiene en cuenta la importancia de cada valor para el total global. Tomemos, por ejemplo, el caso de la compañía de la tabla 3-10, que usa tres grados de mano de obra (no cualificada, semicualificada y cualificada) para fabricar dos productos finales. La compañía quiere conocer el costo promedio de la mano de obra por hora de cada uno de los productos.

TABLA 3-10 Horas de trabajo en el proceso de fabricación

		Horas de trabajo por unidad de producción	
GRADO DE TRABAJO	SUELDO POR HORA (x)	PRODUCTO 1	PRODUCTO 2
No cualificado	$4.00	1	4
Semicualificado	6.00	2	3
Cualificado	8.00	5	3

Un simple promedio aritmético de las tarifas de mano de obra será:

$$\bar{x} = \frac{\Sigma x}{n} \qquad [3\text{-}2]$$

$$= \frac{\$4 + \$6 + \$8}{3}$$

$$= \frac{\$18}{3}$$

$$= \$6.00/\text{hora}$$

En este caso, la media aritmética es incorrecta

Usando esta tasa promedio, podríamos calcular el costo de la mano de obra de una unidad del producto 1 y vemos que es $6(1 + 2 + 5) = $48 y el de una unidad del producto 2 será $6(4 + 3 + 3) = $60. Pero estas respuestas son incorrectas.

Para que sean correctas, han de tener en cuenta el hecho de que se recurre a diferentes cantidades de cada grado. Podemos determinar las respuestas correctas en la siguiente forma. Para el producto 1, el costo total de la mano de obra por unidad es ($4 × 1) + ($6 × 2) + ($8 × 5) = $56, y como hay ocho horas de mano de obra, el costo promedio de ella por hora será $56/8 = $7.00 por hora. Para el producto 2, el costo total de mano de obra por unidad es ($4 × 4) + ($6 × 3) + ($8 × 3) = $58 en el caso del costo promedio de mano de obra por hora de $58/10 o $5.80 por hora.

La respuesta correcta es la media ponderada

Otra manera de calcular el costo promedio correcto por hora para dos productos es tomar el *promedio ponderado* del costo de los tres grados de la mano de obra. Para hacerlo, asignamos un valor al sueldo por hora de cada grado por su proporción del trabajo total que se requiere para fabricar el producto. Una unidad del producto 1, por ejemplo, necesita ocho horas de trabajo. La mano de obra no cualificada usa 1/8 de ese tiempo, la mano de obra semicualificada usa 2/8 de ese tiempo y la mano de obra cualificada necesita 5/8 del tiempo. Si utilizamos esas fracciones como nuestros pesos, una hora de mano de obra para el producto 1 cuesta en promedio:

$$\left(\frac{1}{8} \times \$4\right) + \left(\frac{2}{8} \times \$6\right) + \left(\frac{5}{8} \times \$8\right) = \$7.00/\text{hora}$$

De manera análoga, una unidad del producto 2 requiere diez horas de trabajo, de las cuales 4/10 se destinan a la mano de obra no cualificada, 3/10 a la mano de obra semicualificada y 3/10 a la mano de obra cualificada. Usando como pesos estas fracciones, una hora de mano de obra para el producto 2 cuesta:

$$\left(\frac{4}{10} \times \$4\right) + \left(\frac{3}{10} \times \$6\right) + \left(\frac{3}{10} \times \$8\right) = \$5.80/\text{hora}$$

Cálculo de la media ponderada

Así pues, observamos que los promedios ponderados dan los valores correctos de los costos promedio de la mano de obra por hora para dos productos, pues tienen

en cuenta el hecho de que en éstos se usan diferentes cantidades de cada grado de mano de obra.

En forma simbólica, la fórmula con que se calcula el promedio ponderado es:

$$\bar{x}_w = \frac{\Sigma(w \times x)}{\Sigma w} \qquad [3\text{-}5]$$

donde

- $\bar{x}_w$ = símbolo de la media ponderada*
- w = peso asignado a cada observación (1/8, 2/8, y 5/8 para el producto 1 en nuestro ejemplo)
- $\Sigma(w \times x)$ = suma del peso de cada elemento multiplicado por ese elemento
- Σw = suma de todos los pesos

Si aplicamos la ecuación 3-5 al producto 1 en el ejemplo del costo de la mano de obra, obtenemos:

$$\bar{x}_w = \frac{\Sigma(w \times x)}{\Sigma w} \qquad [3\text{-}5]$$

$$= \frac{(1/8 \times \$4) + (2/8 \times \$6) + (5/8 \times \$8)}{1/8 + 2/8 + 5/8}$$

$$= \frac{\$7}{1}$$

$$= \$7.00/\text{hora}$$

Media aritmética de datos agrupados: la media ponderada

Adviértase que la ecuación 3-5 establece de modo más formal algo que ya hemos hecho antes. Al calcular la media aritmética a partir de datos agrupados (página 71), en realidad obtuvimos la media ponderada utilizando las marcas de clase para los valores x y las frecuencias de cada clase como los pesos. Este producto lo dividimos entre la suma de todas las frecuencias, que fue lo mismo que dividirla entre la suma de todos los pesos.

De manera semejante, *cualquier* media calculada de todos los valores en un conjunto de datos según la ecuación 3-1 o 3-2 es en realidad un promedio ponderado de los componentes del conjunto de datos. Así, en una fábrica podríamos determinarlo a partir de los sueldos (de trabajadores cualificados, semicualificados o no cualificados) o bien los de los hombres, mujeres o de los miembros sindicalizados y no sindicalizados.

* El símbolo $\bar{x}_w$ se lee *x barras bajo w*. La w minúscula se llama subíndice y es un recordatorio de que ésta no es una media ordinaria, sino que está ponderada conforme a la importancia relativa que tienen los valores de x.

EJERCICIOS

3-16 Un profesor ha decidido utilizar un promedio ponderado al calcular las calificaciones finales de los estudiantes que asistieron a su seminario. El promedio de las tareas hechas en casa representará 20% de cada calificación; el examen parcial, 25%; el examen final, 35%; el examen trimestral, 10%; y los problemas, 10%. Con los datos anexos calcule el promedio final de los cinco estudiantes que asistieron al seminario.

ALUMNO	TAREA ESCOLAR	PROBLEMAS	EX. TRIMESTRAL	EX. PARC.	EX. FINAL
1	85	89	94	87	90
2	78	84	88	91	92
3	94	88	93	86	89
4	82	79	88	84	93
5	95	90	92	82	88

3-17 Una tienda anuncia: "Si nuestros precios promedio no son iguales o más bajos que los de cualquier otra tienda, le regalamos la mercancía." Uno de los clientes llega un día a la tienda y pone sobre el mostrador facturas de 6 artículos que compró a un competidor pagando un precio promedio menor. He aquí el importe de las facturas:

$1.29 $2.97 $3.49 $5.00 $7.50 $10.95

Los precios que esas mismas 6 mercancías tenían en la tienda son: $1.35, $2.89, $3.19, $4.98, $7.59 y $11.50. El dueño dijo al cliente: "Mi anuncio se refiere al promedio ponderado del precio de esos artículos. Nuestro promedio es menor porque nuestras ventas de ellos han sido:"

7 9 12 8 6 3

¿Está el dueño de la tienda metiéndose en problemas o saliendo de ellos al hablar de promedios ponderados?

3-18 Una compañía mueblera publicó 6 anuncios en los diarios locales durante diciembre. Se produjo la siguiente distribución de frecuencia:

Número de veces que los suscriptores vieron el anuncio en diciembre	0	1	2	3	4	5	6
Frecuencia	897	1,082	1,325	814	307	253	198

¿Cuál es el número promedio de veces que un suscriptor vio un anuncio de esta compañía en diciembre?

3-19 La sucursal de un gran fabricante de aparatos domésticos está haciendo los pronósticos de sus ventas regionales para el siguiente año. La sucursal del Atlántico, cuyas ventas anuales ascienden en este momento a $193.8 millones, deberá alcanzar un crecimiento de ventas de 7.25%; la sucursal del oeste medio, cuyas ventas son actualmente de $79.3 millones, debe crecer 8.20%; y la sucursal del Pacífico, con ventas de $57.5 millones, deberá crecer en 7.15%. ¿Cuál es la tasa promedio del crecimiento de ventas que se pronostica para el siguiente año?

3-20 El U.S. Postal Service maneja 7 tipos básicos de cartas y tarjetas: tercera clase, segunda clase, primera clase, correo aéreo, entrega inmediata, correo registrado y correo certificado. El volumen de la correspondencia durante 1977 se da en la siguiente tabla:

TIPO DE CORREO	ONZAS ENTREGADAS (EN MILLONES)	PRECIO POR ONZA
Tercera clase	16,400	$.05
Segunda clase	24,100	.08
Primera clase	77,600	.13
Correo aéreo	1,900	.17
Entrega inmediata	1,300	.35
Correspondencia registrada	750	.40
Correspondencia certificada	800	.45

¿Cuál fue el ingreso promedio por onza de estos servicios durante el año?

3-21 Una firma de consultoría administrativa tiene 4 tipos de profesionales en su personal: consultores administrativos, socios de alto rango, personal de campo y personal de oficina. Las tarifas promedio que se cobran a los clientes por el trabajo de cada una de esas categorías profesionales es $75/hora, $40/hora, $30/hora y $15/hora. Los registros de la oficina indican el siguiente número de horas facturadas el último año en cada categoría: 8,000, 14,000, 24,000 y 35,000. Si la empresa está tratando de encontrar una tasa promedio de facturación para estimar las cuentas de los clientes para el próximo año, ¿qué recomienda que haga y cuál es, a su juicio, la tarifa adecuada?

3-4 UNA TERCERA MEDIDA DE TENDENCIA CENTRAL: la media geométrica

Cálculo de la tasa de crecimiento: la media geométrica

Algunas veces estamos manejando cantidades que cambian a lo largo de un periodo; entonces necesitamos conocer una tasa promedio de cambio, como el crecimiento promedio a través de un periodo de varios años. En tales casos, la simple media aritmética no es apropiada, porque no proporciona las respuestas correctas. Lo que necesitamos encontrar es la *media geométrica*, a veces abreviada simplemente como M.G.

Consideremos, por ejemplo, el crecimiento de una cuenta de ahorros. Supóngase que depositamos $100 inicialmente y dejamos que el interés se acumule a

TABLA 3-11 Crecimiento de depósitos de $100 en una cuenta de ahorros

AÑO	TASA DE INTERES	FACTOR DE CRECIMIENTO	AHORROS AL FINAL DEL AÑO
1	7%	1.07	$107.00
2	8	1.08	115.56
3	10	1.10	127.12
4	12	1.12	142.37
5	18	1.18	168.00

diferentes tasas por cinco años. El crecimiento se resume en la tabla 3-11. El rótulo "factor de crecimiento" es igual a:

$$1 + \frac{\text{tasa de interés}}{100}$$

El factor de crecimiento es la cantidad por la cual multiplicamos los ahorros al inicio del año para obtener los ahorros al final del año. Este simple factor de crecimiento de la media aritmética será (1.07 + 1.08 + 1.10 + 1.12 + 1.18)/5 = 1.11, que corresponde a una tasa promedio de interés de 11% al año. Si el banco rinde intereses a una tasa constante de 11% anuales, un depósito de $100 crecerá en cinco años en:

En este caso, la media aritmética de la tasa de crecimiento es incorrecta

$$\$100 \times 1.11 \times 1.11 \times 1.11 \times 1.11 \times 1.11 = \$168.51$$

La tabla 3-11 muestra que la cifra real es apenas $168.00. Así pues, el factor correcto de crecimiento promedio será ligeramente menor que 1.11.

Cálculo de la media geométrica

Para encontrar el factor correcto del crecimiento promedio, multiplicaremos juntos los factores de crecimiento de cinco años y luego sacaremos la quinta raíz del producto, o sea el número que al ser multiplicado por sí mismo es igual al producto inicial. El resultado es la tasa de crecimiento de la *media geométrica*, la cual es el promedio que hemos de utilizar aquí. La fórmula con que se obtiene la media geométrica de una serie de números es:

$$\boxed{\text{Número de valores } x} \quad \text{M.G.} = \sqrt[n]{\text{Producto de todos los valores } x} \qquad [3\text{-}6]$$

Si aplicamos la ecuación anterior al problema de la cuenta de ahorros, podremos determinar que 1.1093 es el factor correcto del crecimiento promedio.

$$\begin{aligned} \text{M.G.} &= \sqrt[n]{\text{Producto de todos los valores } x} \qquad [3\text{-}6] \\ &= \sqrt[5]{1.07 \times 1.08 \times 1.10 \times 1.12 \times 1.18} \\ &= \sqrt[5]{1.679965} \\ &= 1.1093 \leftarrow \text{factor de crecimiento promedio} \end{aligned}$$

Advertencia: use la media apropiada

Nótese que la tasa correcta del interés promedio de 10.93% al año, obtenida mediante la media geométrica, se acerca mucho a la tasa promedio incorrecta de 11% calculada con la media aritmética. Ello se debe a que las tasas de interés son relativamente pequeñas. Pero no sucumbamos a la tentación de recurrir a la media aritmética en vez de a la media geométrica, la cual es más complicada. La razón de ello se aprecia en el siguiente ejemplo.

En las economías muy inflacionarias, los bancos han de pagar altas tasas de interés para captar ahorros. Supóngase que, durante un periodo de cinco años, en una economía extraordinariamente inflacionaria los bancos pagan intereses a las tasas anuales de 100, 200, 250, 300 y 400%, que corresponden a los factores de crecimiento de 2, 3, 3.5, 4 y 5. (Hemos calculado dichos factores tal como lo hicimos en la Tabla 3-11.)

En un lapso de cinco años, un depósito inicial de \$100 crecerá a \$100 × 2 × 3.5 × 4 × 5 = \$42,000. El factor de crecimiento de la media aritmética es (2 + 3 + 3.5 + 4 + 5)/5, o sea 3.5. Esto corresponde a una tasa promedio de interés de 250%. No obstante, si los bancos en realidad dan intereses a una tasa constante de 250% al año, entonces \$100 llegarán a \$52,521.88 en cinco años.

$$\$100 \times 3.5 \times 3.5 \times 3.5 \times 3.5 \times 3.5 = \$52{,}521.88$$

Esta respuesta rebasa la respuesta real de \$42,000 en más de \$10,500, lo cual es un error considerable.

Usemos ahora la fórmula para obtener la media geométrica de una serie de números que nos permitan determinar el factor correcto de crecimiento:

$$\begin{aligned} \text{M.G.} &= \sqrt[n]{\text{Producto de todos los valores } x} \qquad [3\text{-}6]\\ &= \sqrt[5]{2 \times 3 \times 3.5 \times 4 \times 5}\\ &= \sqrt[5]{420}\\ &= 3.347 \leftarrow \text{factor de crecimiento promedio} \end{aligned}$$

Este factor de crecimiento corresponde a una tasa promedio de interés de 235% al año. En este caso, el empleo de la media apropiada *sí* resulta decisivo.

EJERCICIOS

3-22 Una empresa textil ha mostrado los siguientes incrementos porcentuales en el capital durante los últimos 5 años.

1982	1983	1984	1985	1986
5%	10.5%	9.0%	6.0%	7.5%

¿Cuál es el incremento porcentual promedio en el capital durante el periodo de cinco años?

3-23 A continuación se incluye el crecimiento en los gastos que una empresa ha hecho en los últimos años por concepto de cuentas incobrables. Calcule el incremento porcentual promedio en ese gasto durante este periodo. Si esa tendencia continúa, estime el incremento porcentual en cuentas incobrables para 1988.

1980	1981	1982	1983	1984	1985	1986
.11	.09	.075	.08	.095	.108	.120

3-24 Un fabricante de tarjetas de circuitos eléctricos ha producido el siguiente número de unidades en los últimos 5 años:

1982	1983	1984	1985	1986
12,500	13,250	14,310	15,741	17,630

Calcule el incremento porcentual promedio en unidades producidas durante este periodo, y con ese dato estime la producción de 1989.

3-25 Juan Torres está calculando el factor de crecimiento promedio de su tienda de equipo de estéreo en los últimos 6 años. Por medio de la media geométrica encuentre una respuesta de 1.24. Los factores individuales de crecimiento en los primeros años fueron 1.19, 1.35, 1.19 y 1.30, pero Juan perdió los registros del sexto año luego de haber calculado la media. ¿Cuál fue el factor de ese año?

3-26 A través de un periodo de 3 semanas, el propietario de una tienda compró \$120 de hojas de acrílico para las nuevas cajas de exhibición en tres compras iguales de \$40 cada una. La primera compra fue de \$1.00 por pie cuadrado; la segunda, \$1.10; y la tercera, \$1.15. ¿Cuál fue el precio promedio por pie cuadrado pagado por todas las hojas?

3-27 Una tienda de aparatos electrónicos pone a su mercancía un sobreprecio de 35% por encima del costo de sus últimas adiciones al inventario. Hasta hace 4 meses, la grabadora Dynamic 400-S VHS costaba \$300. Durante los últimos 4 meses la empresa ha recibido 4 embarques mensuales de esta grabadora a los siguientes costos unitarios: \$275, \$250 y \$225. ¿A qué tasa promedio mensual ha ido aumentando el precio al menudeo durante este periodo de 4 meses?

3-28 Una proveedora industrial tiene registros del costo del procesamiento de una orden de compra. En los últimos 5 años, el costo ha mostrado la siguiente tendencia: \$55,00, \$58,00, \$61,00, \$65,00, y \$66,00? ¿Cuál ha sido, durante este periodo, el porcentaje promedio de incremento de la industria? Si esa tasa promedio permanece inalterada más de 3 años, ¿cuánto costará a la industria procesar una orden de compra en ese tiempo?

3-29 Un sociólogo ha estado estudiando los cambios anuales en el número de convictos enviados a los centros de readaptación más grande del estado. Sus datos se expresan en función del incremento porcentual del número de prisioneros (un número negativo indica una disminución porcentual). A continuación se transcriben los datos más recientes del sociólogo:

1981	1982	1983	1984	1985	1986
−4%	5%	10%	3%	6%	−5%

a) Calcule el incremento porcentual empleando los datos de 1982 a 1985.
b) Reelabore la parte (a) utilizando los datos de los 6 años.
c) En 1980 se aprobó un nuevo código penal. Antes, la población de las cárceles crecía a una tasa de 2% al año. ¿Cuál parece ser el efecto del nuevo código?

3-5 UNA CUARTA MEDIDA DE TENDENCIA CENTRAL: la mediana

Definición de la mediana

La mediana es una medida de tendencia central diferente de las medias que hemos venido explicando hasta ahora. La mediana es un solo valor del conjunto de datos que mide el elemento central en los datos. Ese elemento es el *más central* en el conjunto de números. La mitad de los elementos se encuentran por arriba de este punto y la otra mitad cae debajo de él.

Cálculo de la mediana a partir de datos no agrupados

Cálculo de la mediana de datos no agrupados

Al calcular la mediana de un conjunto de datos, primero se organizan en un arreglo por orden descendente o ascendente. Si el conjunto contiene un número *impar* de elementos, el de la mitad del arreglo será la mediana. Si existe un número *par* de elementos, la mediana será el promedio de los dos que se hallen en la mitad. En lenguaje formal, la mediana es:

$$\text{Mediana} = \text{el } \left(\underbrace{\frac{n+1}{2}}_{\text{Número de elementos del arreglo}}\right)\text{-ésimo elemento en un arreglo de datos} \qquad [3\text{-}7]$$

Número impar de elementos

Supóngase que queremos encontrar la mediana de varios elementos en un arreglo de datos. Conforme a la ecuación 3-7, la mediana será el $(7 + 1)/2 = 4$o. elemento en el arreglo. Si aplicamos esto a nuestro ejemplo anterior de los tiempos de siete miembros de un equipo de pista, descubriremos que el cuarto elemento del arreglo es 4.8 minutos. Este es el tiempo medio del equipo. Adviértase que, a diferencia de la media aritmética que calculamos antes, la mediana que obtuvimos en la tabla 3-12 *no* fue deformada por la presencia del último valor (9.0). Este valor pudo haber sido 15.0 e incluso 45.0 minutos, y la mediana húbiera sido igual.

Mediana no distorsionada por valores extremos

TABLA 3-12 Tiempos de los miembros del equipo de pista

ELEMENTO EN EL ARREGLO DE DATOS	1	2	3	4	5	6	7
TIEMPO EN MINUTOS	4.2	4.3	4.7	4.8	5.0	5.1	9.0
				↑ mediana			

Número par de elementos

Y ahora calculemos la mediana de un arreglo con un número par de elementos. Examinemos los datos que aparecen en la tabla 3-13 concernientes al número de pacientes atendidos diariamente en la sala de urgencias de un hospital. Los datos se disponen por orden descendente. Su mediana será:

$$\text{Mediana} = \text{el } \left(\frac{n+1}{2}\right)\text{-ésimo elemento en un arreglo de datos} \qquad [3\text{-}7]$$

$$= \frac{8+1}{2}$$

$$= 4.5\text{-ésimo elemento}$$

Dado que la mediana es el 4.5-ésimo elemento del arreglo, necesitamos promediar los elementos cuarto y quinto. El cuarto elemento en la tabla 3-13 es 43 y el quinto es 35. El promedio de estos dos elementos es igual a $(43 + 35)/2$, o sea 39. Por tanto, esta última cifra es el número medio de pacientes atendidos en la sala de urgencias por día durante un periodo de 8 días.

TABLA 3-13 Pacientes atendidos en la sala de urgencias durante 8 días consecutivos

ELEMENTO EN EL ARREGLO DE DATOS	1	2	3	4	5	6	7	8
NUMERO DE PACIENTES	86	52	49	43	35	31	30	11
				↑ mediana de 39				

Cálculo de la mediana a partir de datos agrupados

Cálculo de la mediana de datos agrupados

A menudo tenemos acceso a los datos sólo después de haberlos agrupado en una distribución de frecuencia. Por ejemplo, no conocemos todas las observaciones que han llevado a construir la tabla 13-4, es decir, los datos sobre 600 clientes del banco a quienes nos hemos referido en páginas precedentes. Por el contrario, tenemos diez intervalos de clase y un registro de la frecuencia con que la observación aparece en cada intervalo.

TABLA 3-14 Saldos mensuales promedio de 600 clientes

CLASE EN DOLARES	FRECUENCIA	
0– 49.99	78	
50.00– 99.99	123	
100.00–149.99	187	← mediana de clase
150.00–199.99	82	
200.00–249.99	51	
250.00–299.99	47	
300.00–349.99	13	
350.00–399.99	9	
400.00–449.99	6	
450.00–499.99	4	
	600	

Localización de la mediana de clase

Pese a ello, podemos calcular el saldo medio de la cuenta de cheques de esos 600 clientes al determinar cuál de los diez intervalos de clase *contiene* a la mediana. Para hacerlo es preciso sumar las frecuencias en la columna correspondiente a ellas en la tabla 3-14 hasta llegar al $(n + 1)$ 2o. elemento. Puesto que hay 600 cuentas, el valor de $(n + 1)/2$ es 300.5 (o sea el promedio del 300-ésimo y 301-ésimo elementos). El problema reside en encontrar los intervalos de clase que contienen a esos elementos. La frecuencia acumulativa de las dos primeras clases es apenas $78 + 123 = 201$. Pero cuando llegamos al tercer intervalo, se suman 187 elementos a 201 dándonos un total de 388. Por tanto, las observaciones 300 y 301 deben situarse en esta tercera clase (el intervalo entre \$100.00 y \$149.99).

Interpolación para obtener la mediana

La *mediana de clase* de este conjunto de datos contiene 187 elementos. Si suponemos que éstos comienzan en \$100.00 y que tienen un *espaciamiento uniforme en todo el intervalo de clase* de \$100.00 a \$149.00, podremos interpolar y obtener los valores de los elementos 300 y 301. Primero determinamos que el elemento 300 es el nonagésimo noveno elemento en la mediana de clase:

$$300 - 201 \text{ [elementos en las dos primeras clases]} = 99$$

y que el elemento 301 es el elemento centésimo de la mediana de clase

$$301 - 201 = 100$$

Después calculamos la *amplitud* de los 187 pasos iguales de \$100.00 a \$149.99 como sigue:

Primer elemento de la clase siguiente

Primer elemento de la mediana de clase

$$\frac{\$150.00 - \$100.00}{187} = \$.267 \text{ de amplitud}$$

Ahora bien, si hay 187 pasos de \$.267 cada uno y si 98 pasos nos llevarán al nonagésimo noveno elemento, entonces este elemento será:

$$(\$.267 \times 98) + \$100 = \$126.17$$

y el elemento 100 es un paso adicional:

$$\$126.17 + \$.267 = \$126.44$$

Por tanto, podemos utilizar \$126.17 y \$126.44 como los valores de los elementos 300 y 301, respectivamente.

La mediana real de este conjunto de datos es el valor del elemento 300.5-ésimo; es decir, el promedio de los elementos 300 y 301. Ese promedio será:

$$\frac{\$126.17 + \$126.44}{2} = \$126.30$$

La cifra anterior (\$126.30) es la mediana mensual del saldo de la cuenta de cheques, tal como se estimó con los datos agrupados de la tabla 3.14.

En resumen, podemos calcular del modo siguiente la mediana de los datos agrupados:

Pasos que se siguen para obtener la mediana de datos agrupados

1. Use la ecuación 3-7 para determinar cuál elemento en la distribución se halla más al centro (en este caso, el promedio de los elementos 300 y 301).
2. Sume las frecuencias de cada clase para obtener la que contenga al elemento situado más al centro (la tercera clase, o sea \$100.00 – \$149.99).
3. Determine el número de elementos en la clase (187) y la localización en la clase del elemento de la mediana (el elemento 300 fue el elemento nonagésimo noveno; el elemento 301, el elemento número cien).
4. Encuentre la amplitud de cada paso en la mediana de clase al dividir el intervalo de clase entre el número de elementos presentes en la clase (amplitud = \$.267).
5. Determine el número de pasos entre el límite inferior, la mediana de clase y el elemento correspondiente de la mediana (98 pasos para el elemento nonagésimo noveno; 99 pasos por el elemento número cien).
6. Calcule el valor estimado del elemento de la mediana multiplicando el número de pasos del elemento de la mediana por la amplitud de cada uno y sumando el resultado al límite inferior de la clase de la mediana (\$100 + 98 × \$.267 = 126.17; \$126.17 + \$.267 = \$126.44).
7. Si, como en el ejemplo anterior, hay un número impar de elementos en la distribución, promedie los valores del elemento de la mediana calculado en el paso # 6 (\$126.30).

Un método más fácil

Para abreviar este procedimiento, los estadísticos aplican una ecuación que les permita determinar la mediana de los datos agrupados. En el caso de una muestra, dicha ecuación será:

Mediana muestral

$$\tilde{m} = \left(\frac{(n+1)/2 - (F+1)}{f_m}\right) w + L_m \qquad [3\text{-}8]$$

donde:

- $\tilde{m}$ = mediana de la muestra
- n = número total de elementos de la distribución
- F = suma de todas las frecuencias de clase *hasta* la mediana de clase, pero sin *incluirla*.
- f_m = frecuencia de la mediana de clase
- w = amplitud de intervalo de las clases
- L_m = límite inferior del intervalo de mediana de clase

Si aplicamos la ecuación 3-8 para calcular la mediana de la muestra de saldos de las cuentas de cheques, entonces $n = 600$, $F = 201$, $f_m = 187$, $w = \$50$ y $L_m = \$100$.

$$\tilde{m} = \left(\frac{(n+1)/2 - (F+1)}{f_m}\right) w + L_m \qquad [3\text{-}8]$$
$$= \left(\frac{601/2 - 202}{187}\right) \$50 + \$100$$
$$= \left(\frac{300.5 - 202}{187}\right) \$50 + \$100$$
$$= \left(\frac{98.5}{187}\right) \$50 + \$100$$
$$= (.527)(\$50) + \$100$$
$$= \$126.35 \leftarrow \text{mediana muestral estimada}$$

La ligera diferencia entre esta respuesta y la nuestra, calculada con el método largo, se debe al redondeo.

Ventajas y desventajas de la mediana

Ventajas de la mediana

La mediana ofrece varias ventajas sobre la media. La más importante, demostrada en el ejemplo del equipo de pista en la tabla 3-12, consiste en que los valores extremos no la afectan tan profundamente como a la media. La mediana es fácil de entender y puede ser calculada con cualquier clase de datos (aun a partir de datos agrupados con clases abiertas, como la distribución de frecuencia en la tabla 3-9), a *menos* que la mediana caiga dentro de una clase abierta.

Podemos obtener la mediana aun cuando los datos sean descripciones cualitativas, como el color o la nitidez, en vez de números. Por ejemplo, supóngase que tenemos cinco corridas de una impresora, cuyos resultados han de clasificarse de acuerdo con la nitidez de la imagen. Podemos ordenar los resultados desde los mejores hasta los peores: extremadamente nítidos, muy nítidos, nítidos, ligeramente borrosos, borrosos y muy borrosos. La mediana de las cinco clasificaciones es $(5 + 1)/2$, o sea la tercera clasificación (nítidos).

Desventajas de la mediana

La mediana tiene también desventajas. Ciertos procedimientos estadísticos que se sirven de ella son más complejos que los que usan la media. Por lo demás, como la mediana es un promedio de posición, debemos organizar los datos antes de realizar los cálculos. Se trata de un proceso lento para cualquier conjunto de datos que tenga un vasto número de elementos. En consecuencia, si queremos utilizar un estadístico muestral como estimación de un parámetro de la población, la media es más fácil de aplicar que la mediana. En el capítulo 8 hablaremos más detenidamente de la estimación.

EJERCICIOS

3-30 Una compañía de transportes conserva los registros del kilometraje en todo su equipo rodante. A continuación se anotan los registros del kilometraje semanal de sus camiones.

810	450	756	789	210	657	589	488	876	689
1,450	560	469	890	987	559	788	943	447	775

a) Calcule la mediana de kilómetros que recorre un camión.
b) Calcule la media de 20 camiones
c) Compare (a) y (b) y explique cuál es la mejor medida de tendencia central de los datos.

3-31 Una cadena de supermercados compara los precios por mercancías idénticas en todas sus tiendas de comestibles. En seguida se transcriben los precios a que una libra de tocino se vendió en cada tienda la semana anterior:

$1.08	.98	1.09	1.24	1.33	1.14	1.55	1.08	1.22	1.05

a) Calcule el precio de la mediana por libra.
b) Calcule el precio medio por libra
c) ¿Qué valor es la mejor medida de tendencia central de estos datos?

3-32 Para la distribución de frecuencia que se indica a continuación, determine:
a) Cuál es la mediana de clase.
b) Cuál número de elemento representa la mediana del elemento.
c) La amplitud de los pasos iguales en la mediana de clase.
d) El valor estimado de la mediana de éstos datos.

CLASE	FRECUENCIA	CLASE	FRECUENCIA
100–149.5	12	300–349.5	72
150–199.5	14	350–399.5	63
200–249.5	27	400–449.5	36
250–299.5	58	450–499.5	18

3-33 Los siguientes datos representan los pesos de peces atrapados en un barco:

CLASE	FRECUENCIA
0– 24.9	5
25– 49.9	13
50– 74.9	16
75– 99.9	8
100–124.9	6

a) Con la ecuación 3-8 estime la mediana del peso de los peces atrapados.
b) Con la ecuación 3-3 calcule la media de estos datos.
c) Compare (a) y (b) y diga cuál es la mejor medida de tendencia central de estos datos.

3-34 La compañía Chicago Transit Authority piensa que la excesiva velocidad de sus autobuses incrementa el costo del mantenimiento. En su opinión, un tiempo medio razonable del O'Hare

Airport al John Hancock Center sería de unos 30 minutos. ¿Pueden los datos de la muestra adjunta (en minutos) ayudarle a determinar si los autobuses han sido conducidos a velocidades demasiado altas? Si su conclusión de esos datos es afirmativa, ¿qué explicación recibiría de los conductores?

17	32	21	22
29	19	29	34
33	22	28	33
52	29	43	39
44	34	30	41

3-35 El gerente de una compañía está investigando la cantidad de material que utiliza su empresa en los trabajos de tapicería. La cantidad varía según los trabajos, debido a los distintos estilos de mobiliario y a los tamaños. El gerente reúne los siguientes datos (en yardas) de los trabajos terminados la semana anterior. Calcule la mediana de yardas usadas en un trabajo la semana anterior.

$5\frac{1}{4}$	$6\frac{1}{4}$	6	$7\frac{7}{8}$	$9\frac{1}{4}$	$9\frac{1}{2}$	$10\frac{1}{2}$
$5\frac{3}{8}$	6	$6\frac{1}{4}$	8	$9\frac{1}{2}$	$9\frac{7}{8}$	$10\frac{1}{4}$
$5\frac{1}{2}$	$5\frac{7}{8}$	$6\frac{1}{2}$	$8\frac{1}{4}$	$9\frac{3}{8}$	$10\frac{1}{4}$	$10\frac{1}{8}$
$5\frac{7}{8}$	$5\frac{3}{4}$	7	$8\frac{1}{2}$	$9\frac{1}{8}$	$10\frac{1}{2}$	$10\frac{1}{8}$
6	$5\frac{7}{8}$	$7\frac{1}{2}$	9	$9\frac{1}{4}$	$9\frac{7}{8}$	10

Si estuvieran programados 150 trabajos en las siguientes tres semanas, por medio de la mediana determine cuántas yardas de material se necesitarán.

3-36 Si las reclamaciones del seguro por accidentes automovilísticos se ajustan a la distribución incluida aquí, determine la mediana aplicando el método descrito en la página 90 de este capítulo. Verifique si obtiene la misma respuesta usando la ecuación 3-8.

IMPORTE DE LA RECLAMACION ($)	FRECUENCIA	IMPORTE DE LA RECLAMACION ($)	FRECUENCIA
Menor que 250	52	750–999.99	1,776
250–499.99	337	1000 en adelante	1,492
500–749.99	1,066		

3-37 Un investigador recibe las siguientes respuestas a un enunciado en una encuesta de evaluación: discrepa fuertemente, discrepa ligeramente, discrepa un poco, concuerda, concuerda fuertemente. ¿Cuál es la mediana de las 6 respuestas?

3-6 UNA ULTIMA MEDIDA DE TENDENCIA CENTRAL: la moda

Definición de la moda

La moda es una medida de tendencia central que difiere de la media, pero que se parece un poco a ella porque realmente no se calcula por medio de los procesos ordinarios de la aritmética. La moda es *el valor que más se repite dentro del conjunto de datos.*

Uso limitado de la moda de datos no agrupados

Como en cualquier otro aspecto de la vida, el azar interviene de manera importante en la ordenación de los datos. Algunas veces hace que un elemento no representativo se repita bastante a menudo y sea el valor más frecuente en el conjunto de datos. Por tal razón, rara vez utilizamos como una medida de tendencia central la moda de datos agrupados. Por ejemplo, la tabla 3-15 muestra el número de viajes de reparto que diariamente hizo una planta de concreto. El valor modal es 15 por ocurrir más a menudo que los demás (tres veces). Una moda de 15 implica que la actividad de la planta es mayor de 6.7 (esta cifra es la respuesta que obtendríamos si calculáramos la media). La moda nos indica que 15 es el número más frecuente de viajes, pero no nos permite conocer que la mayor parte de los valores son menores que 10.

TABLA 3-15 Viajes de reparto por día en un periodo de 20 días

VIAJES DISPUESTOS POR ORDEN ASCENDENTE					
0	2	5	7	15	
0	2	5	7	15	← moda
1	4	6	8	15	
1	4	6	12	19	

Cálculo de la clase modal de datos agrupados

A continuación agruparemos los datos precedentes en una distribución de frecuencia, como lo hicimos en la tabla 3-16. Si seleccionamos la clase con el mayor número de observaciones, a la cual llamaremos *clase modal*, podremos seleccionar "4–7" viajes. Esta clase es más representativa de la actividad de la planta que la moda de 15 viajes por día. Por tal razón, siempre que usemos la moda como medida de tendencia central de un conjunto de datos, habremos de calcularla a partir de datos agrupados.

TABLA 3-16 Distribución de frecuencia de los viajes de reparto

CLASE EN EL NUMERO DE VIAJES	0–3	4–7	8–11	12 y más
FRECUENCIA	6	8	1	5
		↑ clase modal		

La moda en distribuciones simétricas y asimétricas (sesgadas)

Examinemos detenidamente las figuras 3-9 a 3-11, cada una de las cuales contiene una distribución de frecuencia. La figura 3-9 es simétrica. La figura 3-10 está sesgada a la derecha y la figura 3-11 lo está a la izquierda.

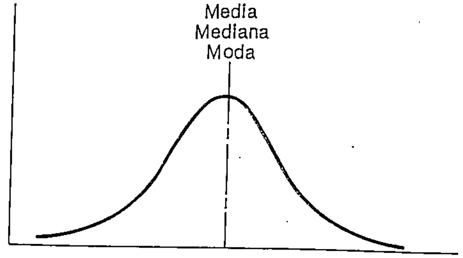


FIGURA 3-9 Distribución simétrica, que muestra que la media, la mediana y la moda coinciden

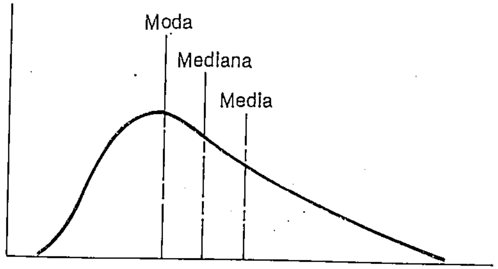


FIGURA 3-10 La distribución está sesgada a la derecha

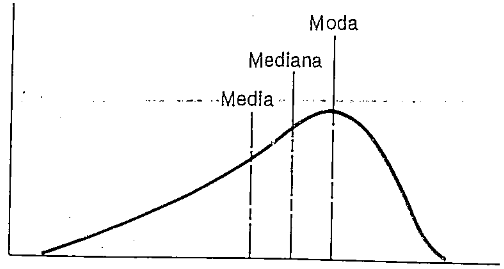


FIGURA 3-11 La distribución está sesgada a la izquierda

Localización de la moda

En la figura 3-9, donde la distribución es simétrica y sólo existe una moda, las tres medidas de tendencia central (la moda, la mediana y la media) coinciden con el punto más alto en la gráfica. En la figura 3-10, el conjunto de datos está sesgado a la derecha. Aquí, la moda está todavía en el punto más alto de la gráfica, pero la mediana está a la derecha de ese punto y la media cae a la derecha de la mediana. Cuando la distribución es asimétrica a la izquierda, como sucede en la figura 3-11, la moda se encuentra en el punto más alto de la gráfica, la mediana se sitúa a la izquierda de la moda y también se halla a la izquierda de la mediana. **Cualquiera que sea la forma de la curva, la moda siempre está situada en el punto más alto.**

Cálculo de la moda a partir de datos agrupados

Cálculo de la moda en la clase modal

Cuando los datos ya están agrupados en una distribución de frecuencia, debemos suponer que la moda se halla en la clase que tenga más elementos; es decir, que posea la frecuencia más elevada. ¿Pero cómo podemos determinar un solo valor de la moda con esta clase modal? Disponemos de dos métodos para ello. El primer método nos permite estimar la moda en una gráfica. El segundo método se sirve de una ecuación.

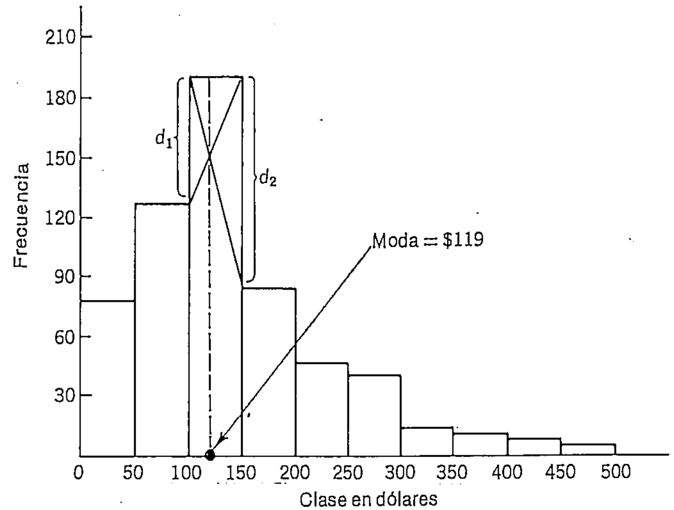


FIGURA 3-12 Cálculo de la moda a partir de los datos agrupados usando el método gráfico

Para demostrar estas dos maneras de calcular la moda en datos agrupados, usemos los datos de la tabla 3-14 en la página 89 (nuestro ejemplo de los saldos de cuentas de cheques). Primero, podemos construir un histograma con los datos como se aprecia en la figura 3-12. Después, por ser la clase modal el rectángulo más alto, encontraremos su moda:

Una solución gráfica

1. Trazando una línea del ángulo superior derecho del rectángulo más alto al angulo superior derecho del rectángulo situado inmediatamente a su izquierda.

2. Trazando una segunda línea del ángulo superior izquierdo del rectángulo más alto al ángulo superior izquierdo del rectángulo situado inmediatamente a su derecha.

3. Trazando una línea perpendicular al eje horizontal por el punto donde se cruzan las líneas dibujadas en los pasos 1 y 2.

El valor del eje horizontal marcado por la línea trazada en el paso 3 se aproximará al valor modal. En este caso, la moda es de 119 aproximadamente.

Una solución matemática

Un segundo método de calcular la moda cuando contamos con datos agrupados consiste en aplicar la ecuación 3-9:

Moda

$$Mo = L_{Mo} + \frac{d_1}{d_1 + d_2} w \quad [3\text{-}9]$$

donde:

- L_{Mo} = límite inferior de la clase modal
- d_1 = frecuencia de la clase modal menos la frecuencia de la clase que se encuentra *inmediatamente debajo de ella*
- d_2 = frecuencia de la clase modal menos la frecuencia de la clase que se halla *inmediatamente encima de ella*
- w = amplitud del intervalo de la clase modal

Si mediante la ecuación 3-9 calculamos la moda de los saldos de las cuentas de cheques, entonces L_{Mo} = \$100, d_1 = 187 − 123 = 64, d_2 = 187 − 82 = 105 y w = \$50.

$$Mo = L_{Mo} + \frac{d_1}{d_1 + d_2} w \qquad [3\text{-}9]$$

$$= \$100 + \frac{64}{64 + 105} \$50$$

$$= \$100 + (.38)(\$50)$$

$$= \$100 + \$19$$

$$= \$119.00 \leftarrow \text{Moda}$$

Nuestra respuesta de \$119 es la estimación de la moda utilizando la gráfica o el método matemático de cálculo.

Distribuciones multimodales

Distribuciones bimodales

¿Qué sucede cuando tenemos dos valores diferentes de manera que *cada uno* aparece mayor número de veces que todos los valores en el conjunto de datos? La tabla 3-17 muestra los errores de facturación durante un periodo de 20 días en la oficina de un hospital. Adviértase que tanto 1 como 4 aparecen el mayor número de veces dentro del conjunto de datos. Cada uno lo hace tres veces. De ahí que esa distribución tenga dos modas y reciba el nombre de *distribución bimodal.*

TABLA 3-17 Errores de facturación por día en un periodo de 20 días

ERRORES DISPUESTOS POR ORDEN ASCENDENTE					
0		2		6	9
0		4		6	9
1		4	← Moda	7	10
1	← Moda	4		8	12
1		5		8	12

En la figura 3-13 hemos graficado los datos de la tabla 3-17. Nótese que hay *dos* puntos más altos en la gráfica. Ocurren en los valores de los errores 1 y 4 de facturación. La distribución de la figura 3-14 se llama también bimodal, aun cuando los dos puntos más altos no sean iguales. Por supuesto, esos puntos sobresalen por arriba de los valores contiguos debido a la frecuencia con que se observan.

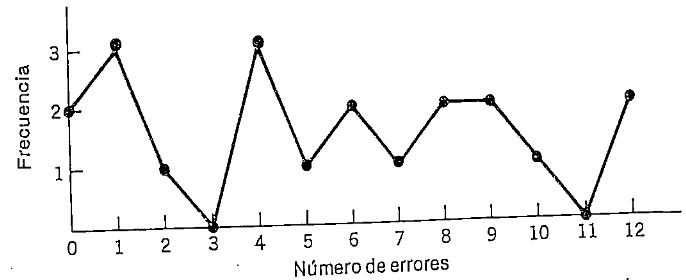


FIGURA 3-13 Datos de la tabla 3-17, que muestra una distribución bimodal

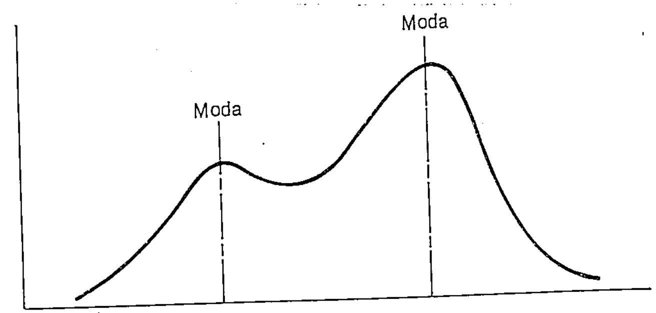


FIGURA 3-14 Distribución bimodal con dos modas desiguales

Ventajas y desventajas de la moda

Ventajas de la moda

La moda, a semejanza de la mediana, puede usarse como una localización central para datos cualitativos y también cuantitativos. Si una impresora produce cinco impresiones que se clasifican en "muy nítidas", "nítidas", "nítidas", "nítidas" y "borrosas", el valor modal será "nítido". De manera análoga, podemos hablar de estilos modales cuando, por ejemplo, los clientes prefieren un mobiliario de estilo tradicional a otro tipo de muebles.

A semejanza de la mediana, **tampoco a la moda le afectan demasiado los valores extremos**. Aun cuando los valores superiores sean muy altos y los valores inferiores sean muy bajos, decidimos que el valor más frecuente del conjunto de datos sea el valor modal. Podemos servirnos de la moda sin importar la magnitud o la dispersión de los valores en el conjunto de datos.

Una tercera ventaja de la moda consiste en que podemos emplearla cuando una o más clases son abiertas. Nótese, entre otras cosas, que la tabla 3-16 en la página 95 contiene la clase abierta "12 o más viajes".

Desventajas de la moda

A pesar de esas ventajas, la moda se usa menos para medir la tendencia central que la media y la mediana. Algunas veces no hay un valor modal porque el conjunto de datos no contiene valores que ocurran más de una vez. Otras veces todos los valores son la moda ya que ocurren el mismo número de veces. Sin duda la moda es una medida inútil en tales casos. Otra desventaja estriba en que, cuando los conjuntos de datos contienen dos, tres o muchas modas, son difíciles de interpretar y comparar.

EJERCICIOS

3-38 A continuación se dan las edades en años de los automóviles con que trabajó Village Autohaus la última semana:

5, 6, 3, 6, 11, 7, 9, 10, 2, 4, 10, 6, 2, 1, 5

a) Calcule la moda de este conjunto de datos.
b) Calcule la media del conjunto de datos.
c) Compare (a) y (b) y diga cuál es la mejor medida de tendencia central de los datos.

3-39 Las edades de los residentes de Twin Lakes Retirement Village están descritas por la siguiente distribución:

CLASE	FRECUENCIA
47–51.9	4
52–56.9	9
57–61.9	13
62–66.9	42
67–71.9	39
72–76.9	20
77–81.9	9

a) Estime el valor modal de la distribución, aplicando el método gráfico.
b) Use ahora la ecuación 3-9 para estimar ese mismo valor.
c) Comente las respuestas que obtuvo en las partes (a) y (b).

3-40 ¿Cuáles son los valores modales de las siguientes distribuciones?

a)

Color del cabello	Negro	Castaño	Rojo	Rubio
Frecuencia	11	24	6	18

b)

Tipo de sangre	AB	O	A	B
Frecuencia	4	12	35	16

c)

Día de nacimiento	Lun.	Mart.	Miér.	Juev.	Viern.	Sáb.	Dom.
Frecuencia	22	10	32	17	13	32	14

3-41 Las edades de una muestra de los estudiantes que asisten en este semestre a cierta universidad son:

19	17	15	20	23	41	33	21	18	20
18	33	32	29	24	19	18	20	17	22
55	19	22	25	28	30	44	19	20	39

a) Construya una distribución de frecuencia con los intervalos 15–19, 20–24, 25–29, 30–34 y 35 y de mayor edad.
b) Estime el valor modal usando la ecuación 3-9.
c) Ahora calcule la media de los datos brutos.
d) Compare sus respuestas en (b) y (c) y señale cuál de las dos es la mejor medida de tendencia central de estos datos y por qué.

3-42 Estime la moda de la distribución dada en el ejercicio 3-36.

3-43 El número de sistemas de calefacción solar con que cuenta el público es muy vasto, y sus capacidades térmicas son también muy variadas. A continuación se da una distribución de la capacidad térmica (en días) de 28 sistemas que fueron probados recientemente en los laboratorios de una empresa:

DIAS	FRECUENCIA
0–0.99	2
1–1.99	4
2–2.99	6
3–3.99	7
4–4.99	5
5–5.99	3
6–6.99	1

La empresa sabe que su informe sobre las pruebas circulará en amplios medios y servirá de base a las leyes tributarias sobre aparatos de calefacción solar. Por tanto, quiere que las medidas que utilice reflejen los datos con la mayor precisión posible.
a) Calcule la media de esos datos.
b) Calcule la moda de esos datos.
c) Calcule la mediana de esos datos.
d) Seleccione la respuesta entre (a), (b) y (c) que mejor refleje la tendencia central de los datos de prueba y justifique su elección.

3-44 Eduardo Grijalva es director de la oficina de ayuda financiera a los estudiantes en una universidad. Ha empleado los datos disponibles referentes a los ingresos de verano de todos los estudiantes que han solicitado ayuda a su oficina y con ellos ha preparado la siguiente distribución de frecuencia:

INGRESOS DE VERANO	NUMERO DE ESTUDIANTES
$ 0–$ 499	231
500– 999	304
1,000– 1,499	400
1,500– 1,999	296
2,000– 2,499	123
2,500– 2,999	68
3,000 o más	23

a) Encuentre la clase modal para los datos del señor Grijalva.
b) Con la ecuación 3-9 obtenga la moda para los datos del señor Grijalva.
c) Si la ayuda financiera se concede exclusivamente a los estudiantes cuyos ingresos de verano fueron por lo menos 10% inferiores a los ingresos modales de verano, ¿cuántos estudiantes llenan ese requisito?

3-7 COMPARACION DE LA MEDIA, LA MEDIANA Y LA MODA

La media, la mediana y la moda son idénticas en una distribución simétrica

Cuando resolvemos un problema estadístico, hemos de decidir si utilizamos la media, la mediana o la moda como medida de tendencia central. Las distribuciones simétricas que contienen sólo una moda siempre tienen el mismo valor para esos tres estadísticos, según se mostró en la figura 3-9 de la página 96. En tales casos, no necesariamente debemos escoger la medida de tendencia central, ya que esa elección ha sido hecha previamente.

En una distribución de sesgo positiva (que tiene asimetría a la derecha, como la de la Fig. 3-10), los valores se concentran en el extremo izquierdo del eje horizontal. Aquí, la moda es el punto más alto de la distribución; la mediana se halla a la derecha de él; y la media se encuentra a la derecha de la moda y la mediana. En una distribución de sesgo negativo, como la de la figura 3-11, los valores se concentran en el extremo derecho del eje horizontal. La moda es el punto más alto de la distribución, y la mediana se halla a la izquierda de él. La media está a la izquierda de la moda y de la mediana.

La mediana puede ser la más idónea en las distribuciones sesgadas

Cuando la población tiene sesgo positivo o negativo, la mediana es a menudo la mejor medida de la ubicación, puesto que siempre se encuentra entre la media y la moda. A la mediana no le afecta tanto la frecuencia de ocurrencia de un solo valor como la moda; tampoco es atraída por valores extremos como la media.

Por lo demás, no se cuenta con pautas universales para aplicar la media, mediana o moda como medida de tendencia central de diferentes poblaciones. Cada clase ha de ser juzgada en forma independiente conforme a las pautas que hemos expuesto.

EJERCICIOS

3-45 Cuando la distribución de los datos es simétrica y de forma de campana, la selección de una medida de localización se simplifica considerablemente. ¿Por qué?

3-46 ¿En qué tipo de distribución (de sesgo positivo, de sesgo negativo o simétrica) es:
a) La media menor que la mediana?
b) La moda menor que la media?
c) La mediana menor que la moda?

3-47 Nora Robles, la agresiva gerente de una división de mercadotecnia de una corporación, acaba de salir muy molesta de una reunión con los analistas estadísticos de la compañía. La bombardearon con información estadística referente a la última campaña publicitaria que dirigió y que incluía muchas medidas de tendencia central para varias secciones de las poblaciones meta. Ha ordenado al jefe de la sección de análisis estadístico prepararle en una hora una explicación de todas las medidas de tendencia central. Usted es el jefe de esa sección y deberá elaborar una lista que contenga exactamente un punto fuerte y un punto débil de cada una de las medidas de tendencia central: media, mediana y moda.

3-8 GLOSARIO DEL CAPITULO

CLASE DE LA MEDIANA (MEDIANA DE CLASE) Clase de distribución de frecuencia que contiene el valor medio de un conjunto de datos.

CODIFICACION Método con que se calcula la media de datos agrupados recodificando los valores de las marcas de clase a valores más simples.

CURTOSIS Grado de pico que presenta una distribución de observaciones.

DISTRIBUCION BIMODAL Distribución de observaciones de datos en que dos valores ocurren con mayor frecuencia que el resto de los que integran el conjunto de datos.

ESTADISTICOS Medidas numéricas que describen las características de una muestra.

ESTADISTICOS RESUMIDOS Números individuales que describen ciertas características de un conjunto de datos.

LEPTOCURTICA Dícese de la distribución que tiene un gran pico.

MEDIA Medida de tendencia central que representa el promedio aritmético de un conjunto de observaciones.

MEDIA GEOMETRICA Medida de tendencia central que se usa para medir la tasa promedio de cambio o crecimiento de alguna cantidad; se calcula tomando la *enésima* raíz del producto de n valores que representan el cambio.

MEDIA PONDERADA Promedio que se calcula a fin de tener en cuenta la importancia de cada valor para el total global; es decir, un promedio donde el valor de cada observación se pondera mediante algún índice de su importancia.

MEDIANA Punto medio de un conjunto de datos; una medida de localización que divide los datos en mitades.

MEDIDA DE DISPERSION Aquella que describe cómo las observaciones están esparcidas en un conjunto de datos.

MEDIDA DE TENDENCIA CENTRAL Aquella que indica el valor que cabe esperar de un punto típico o medio de datos.

MESOCURTICA Dícese de la distribución con un pico moderado.

MODA Valor que se repite más veces en un conjunto de datos. Lo representa el punto más alto en la curva de distribución de un conjunto de datos.

PARAMETROS Valores numéricos que describen las características de una población entera, generalmente representados por letras griegas.

PLATICURTICA Dícese de la distribución con un pico ligero.

SESGO (ASIMETRIA) Grado de concentración de una distribución de datos en uno u otro extremo; ausencia de simetría.

SIMETRICO Característica de una distribución donde una mitad es la imagen de espejo de la otra.

3-9 ECUACIONES UTILIZADAS EN EL CAPITULO

[3-1] $$\mu = \frac{\Sigma x}{N}$$ p. 70

La *media aritmética de la población* es igual a la suma de los valores de todos los elementos de la población (Σx), dividida entre el número de elementos de esta última (N):

[3-2] $$\bar{x} = \frac{\Sigma x}{n}$$ p. 70

Para obtener la *media muestral aritmética*, se suman los valores de todos los elementos de la muestra (Σx) y se divide el total entre el número de elementos de la muestra (n).

[3-3] $$\bar{x} = \frac{\Sigma(f \times x)}{n}$$ p. 72

Para calcular la *media muestral aritmética de datos agrupados*, se obtienen las marcas de clase (x) para cada clase de la muestra. Luego se multiplica cada marca de clase por la frecuencia (f) de observaciones de esa clase, se suman (Σ) esos resultados y se dividen entre el número total de observaciones de la muestra (n).

[3-4] $$\bar{x} = x_0 + w\frac{\Sigma(u \times f)}{n}$$ p. 75

Esta fórmula nos permite calcular la *media muestral aritmética de datos agrupados* usando códigos para no tener que trabajar con marcas de clase grandes o inadecuadas. Asigne esos códigos (u) como sigue: dé el valor de cero a la marca de clase de la mitad (denominada x_0), asigne enteros positivos consecutivos a las marcas mayores de clase que x_0 y enteros negativos consecutivos a las marcas de clase más pequeñas. Luego, multiplique el código asignado a cada clase (u) por la frecuencia (f) de observaciones en cada clase y sume (Σ) todos esos productos. Divida este resultado entre el número total de observaciones en la muestra (n), multiplíquelas por la amplitud numérica del intervalo de clase (w) y sume el valor de la marca de clase a la cual se asignó el código cero (x_0).

[3-5] $$\bar{x}_w = \frac{\Sigma(w \times x)}{\Sigma w}$$ p. 82

La *media ponderada*, $\bar{x}_w$, es el promedio que tiene en cuenta la importancia de cada valor para el total global. Podemos calcular este promedio multiplicando el peso, o proporción, de cada elemento (w) por ese elemento (x), sumando los resultados (Σ) y dividiendo esta cantidad entre el total de los pesos (Σw).

[3-6] $$\text{M.G.} = \sqrt[n]{\text{Producto de todos los valores } x}$$ p. 85

La *media geométrica*, abreviada M.G., es adecuada para medir la tasa promedio de cambio (la tasa de crecimiento) en un periodo. En esta ecuación, n es igual al número de valores x a que se refiere el problema.

[3-7] $$\text{Mediana} = \text{el } \left(\frac{n+1}{2}\right)\text{-ésimo elemento de un arreglo de datos}$$ p. 88

donde: n = número de elementos en el arreglo de datos

La *mediana* es un valor individual que mide el elemento central del conjunto de datos. La mitad de éstos se halla arriba de la mediana, la otra mitad debajo de ella. Si el conjunto contiene un número impar de elementos, el elemento de la mitad del arreglo será la mediana. En el caso de un número impar de elementos, la mediana será el promedio de dos elementos de la mitad. Se aplica esta fórmula cuando los datos no están agrupados.

[3-8] $$\tilde{m} = \left(\frac{(n+1)/2-(F+1)}{f_m}\right) w + L_m$$ p. 91

Esta fórmula nos permite obtener la *mediana muestral de datos agrupados*. En ella, n es igual al número total de elementos de la distribución; F es igual a la suma de todas las frecuencias de clase hasta la mediana de clase, pero sin incluirla; f_m es la frecuencia de observaciones en esta clase; w es la amplitud de intervalo de clase; y L_m es el límite inferior del intervalo de la mediana de clase.

[3-9] $$Mo = L_{Mo} + \frac{d_1}{d_1 + d_2} w$$ p. 98

La *moda* es el valor que se repite más veces en el conjunto de datos. Para calcular la *moda de datos agrupados* (simbolizada como Mo), se aplica esta fórmula y se hace que L_{Mo} = límite inferior de la clase modal; d_1 = frecuencia de la clase modal menos la frecuencia de la clase inmediatamente debajo de ella; d_2 = frecuencia de la clase modal menos la frecuencia de la clase inmediatamente encima de ella; y w = la amplitud del intervalo de la clase modal.

3-10 EJERCICIOS DE REPASO

3-48 La tabla siguiente contiene la distribución relativa de las visitas de venta hechas en el mes inmediato anterior a los clientes de una empresa farmacéutica:

Número de visitas de venta	0	1	2	3	4	5 o más
Frecuencia relativa	.21	.18	.38	.19	.03	.01

¿Cuál es la moda de la distribución? ¿Es la moda una buena medida en este caso?

3-49 En un periodo de 4 semanas, una compañía de servicios públicos usó los siguientes kilotons de carbón por día:

LUNES	MARTES	MIERCOLES	JUEVES	VIERNES	SABADO	DOMINGO
23.4	21.8	26.3	21.8	22.6	28.4	27.5
26.4	26.0	24.3	21.5	22.3	25.3	24.9
23.3	24.1	25.4	25.9	22.0	24.0	26.2
22.2	21.9	27.5	27.0	26.1	25.8	24.4

Usando intervalos de clase de 21.00–21.99, 22.00–22.99, etc., encuentre la clase y el valor modales.

3-50 Una agencia publicitaria ha lanzado una campaña para promover un nuevo cereal. Uno de los comerciales de televisión fue proyectado en un mercado de prueba, y se telefoneó a una muestra de espectadores para verificar la retención del mensaje. Los que vieron el anuncio fueron clasificados según el porcentaje de evocación del mensaje, obteniéndose los siguientes resultados:

Recordación (%)	0–9.9	10–19.9	20–29.9	30–39.9	40–49.9	50–59.9	60–69.9	70–79.9
Frecuencia	1	3	2	7	6	10	12	9

Si la agencia ha establecido un promedio de evocación mínima de 50% para todos los anuncios, ¿puede decirse que éste tiene éxito?

3-51 Una cadena opera 4 cafeterías dentro del área de las plantas industriales. Las cafeterías se parecen en su distribución, equipo y menú. Durante el año anterior, la capacidad promedio de clientes por local fue de 440 personas. Este año, al agregar mesas y más empleados, 2 cafeterías aumentaron su capacidad en 20%, lo cual hizo que las 4 cafeterías tengan igual capacidad de clientes. ¿Qué capacidad tenían antes del incremento?

3-52 En una distribución sesgada, la mejor medida de la tendencia central que debe comunicarse es (escoja una):

a) la media
b) la mediana
c) la media geométrica
d) depende de la dirección del sesgo
e) la moda

¿A qué se debe esto?

3-53 Una empresa que realiza encuestas por entrevista para las compañías y otras organizaciones en Nueva York descubre que el tipo de muestreo que desean sus clientes repercute en el costo de las entrevistas. Si un cliente quiere una muestra por área, se requerirá más tiempo y trabajo, ya que hay que llegar a más áreas. La empresa ha dividido la ciudad en sectores y ha fijado la siguiente tarifa de precios por entrevista, según el número de personas a quienes debe entrevistar en un sector:

Docenas de entrevistas/sector	Primeras 10	siguientes 5	siguientes 5	siguientes 10	siguientes 10	siguientes 10	todas las demás
Costo (por docena)	\$14.00	\$13.30	\$12.60	\$12.25	\$11.90	\$11.55	\$11.20

Si una organización que utiliza un plan muestral quiere que 660 (o 55 docenas) sean entrevistados en uno de los sectores, ¿cuál será el costo promedio por entrevista?

3-54 Usando la distribución de frecuencia del ejercicio 3-69:

a) Construya un histograma de frecuencia
b) Determine gráficamente la moda

3-55 Con los datos del ejercicio 3-35:

a) ¿Cuál es la moda?
b) Construya una distribución de frecuencia utilizando clases de 1 yarda de ancho. ¿Cuál es la clase modal?

c) Estime la moda en una distribución de la parte (b) aplicando la ecuación 3-9. Compare su respuesta con la obtenida al emplear el cálculo gráfico.

3-56 Suponga que la distribución de la figura 3-1 representa la distribución de la producción de calor en 3 tipos de hornos solares.

a) En promedio, ¿cuál horno genera más calor?
b) ¿Es posible que los 3 tipos tengan la misma producción de calor?
c) ¿Es posible que el tipo A alcance una mayor producción que el tipo B?
d) ¿Es posible que el tipo C dé una producción menor que el tipo A?

Explique brevemente sus respuestas.

3-57 Suponga que las 2 curvas mostradas en la figura 3-2 representan la distribución de las posibles ganancias conseguidas en 2 series de juegos en un experimento. Suponga además que ambas curvas son simétricas y sus centros (puntos más altos) representan cero ganancias. Las posiciones a la izquierda del centro representan pérdidas y las de la derecha del centro representan ganancias. ¿Cuál conjunto de juegos, A o B, preferirá basándose en la simple intuición?

3-58 Compare y contraste la posición central, el sesgo y la curtosis de las distribuciones de las razones de alumno/profesor para las siguientes instituciones:

a) Universidades de Estados Unidos
b) Las grandes universidades estatales de la Unión Americana
c) Las universidades particulares de Estados Unidos.

3-59 El desplazamiento total de cada barco que utiliza el Canal de Panamá durante un periodo de una semana se recopiló en la distribución de frecuencia que se incluye a continuación:

Desplazamiento general (Miles de tons)	0–2.99	3–5.99	6–8.99	9–11.99	12–14.99	15–17.99	18–20.99
Frecuencia	32	49	67	33	14	5	2

a) Calcule la media muestral de estos datos, aplicando el método de marcas de clase.
b) Verifique que logra el mismo resultado empleando el método de codificación.

3-60 Una agencia publicitaria está evaluando diferentes medios de promoción para uno de sus clientes, un fabricante de productos de aseo destinados a los ancianos. Un evento que se piensa realizar es una carrera anual de automóviles. En seguida se da la distribución de edades del público que asistirá a las carreras:

EDAD	FRECUENCIA
15–20	5
21–26	20
27–32	24
33–38	30
39–44	46
45–50	88
51–56	103
57–62	44
63–68	30

Si se desea realizar los eventos que llegan a la audiencia que tiene una edad promedio de más de 45 años, ¿llena los requisitos esta carrera de autos?

3-61 La Environmental Protection Agency ha establecido los requisitos del promedio de millas por galón para toda la línea de productos de los fabricantes automotrices nacionales y extranjeros.

La siguiente tabla contiene la distribución de las clasificaciones de millas por galón (MPG) de los automóviles producidos por un fabricante.

MPG	FRECUENCIA RELATIVA
10–12.99	.08
13–15.99	.15
16–18.99	.32
19–21.99	.38
22–24.99	.07

Si el millaje mínimo promedio fijado por esta oficina debe ser 18.5 mpg, ¿cumple con los requisitos este fabricante automotriz?

3-62 Juana Onofre, supervisora del departamento de solicitudes por teléfono, está tratando de reducir los retrasos de los clientes que piden servicio por teléfono. Piensa que con retrasos mínimos disminuirá el malestar de los clientes ante este sistema de pedido. Para definir la magnitud de los retrasos examina una muestra de llamadas telefónicas reuniendo los siguientes datos:

TIEMPO DE RETRASO (EN MINUTOS)	FRECUENCIA	TIEMPO DE RETRASO (EN MINUTOS)	FRECUENCIA
0.00–0.39	15	2.00–2.39	7
0.40–0.79	43	2.40–2.79	4
0.80–1.19	58	2.80–3.19	3
1.20–1.59	21	3.20–3.59	2
1.60–1.99	9	3.60–3.99	2

Si el tiempo medio de retraso de 1.25 minutos o menos es considerado aceptable por Juana, ¿se necesita tomar alguna otra medida?

3-63 Durante 5 años, una empresa de computación ha elevado su publicidad comercial en un promedio de 5.6% al año. Tales incrementos en los 4 primeros años fueron 6.5, 6.1, 5.2 y 4.4%. ¿Cuál fue el aumento en el quinto año?

3-64 Una encuesta de investigación de mercado hecha a 35 familias para conocer su opinión sobre la calidad de cierto programa de televisión arrojó la siguiente distribución de ratings, en que los números positivos indican opiniones favorables y los números negativos (entre paréntesis) indican opiniones desfavorables.

RATING	FRECUENCIA	RATING	FRECUENCIA
(30.0)–(20.1)	2	0.0– 9.9	9
(20.0)–(10.1)	5	10.0–19.9	6
(10.0)–(0.1)	8	20.0–29.9	5

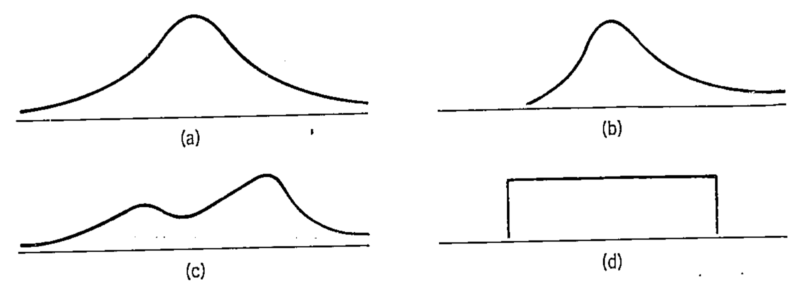
El patrocinador del programa necesita un rating medio de 10.0 por lo menos para continuar patrocinándolo. ¿Cumple el programa con ese requisito?

3-65 El patrón de crecimiento de una comunidad de pensionados en California presenta el siguiente patrón:

AÑO	TASA DE CRECIMIENTO	AÑO	TASA DE CRECIMIENTO
1980	.035	1983	.050
1981	.055	1984	.061
1982	.042	1985	.056

¿Qué tasa media de crecimiento puede anunciar legítimamente la comunidad de pensionados?

3-66 ¿Qué medida de tendencia central recomendaría usted para representar las siguientes distribuciones?



(a) (b) (c) (d)

3-67 Una agencia de relaciones públicas, está preparando el informe anual de una empresa industrial cuyas ventas anuales en los últimos años han sido:

AÑO	VENTAS EN DOLARES (EN MILLONES)
1981	9.8
1982	10.3
1983	11.5
1984	12.6
1985	14.4

¿Qué incremento porcentual promedio en las ventas anuales durante ese periodo puede incluir legítimamente la agencia en su informe anual?

3-68 Los científicos clasifican la radiación atendiendo a la longitud de onda. Las categorías típicas son rayos infrarrojos, visibles, ultravioletas, rayos X y rayos cósmicos. ¿Cuál es la categoría mediana?

3-69 Los pesos de una muestra de paquetes embarcados por avión se dan en la siguiente distribución:

Peso (lbs)	0–9.99	10.0–19.99	20.0–29.99	30.0–39.99	40.0–49.99	50.0 en adelante
Frecuencia	28	25	14	8	4	1

¿Cuál es la mediana? ¿Es una buena medida en este caso?

3-70 El 3 de enero, Anselmo Montes tuvo una jornada mejor de lo habitual en el hipódromo. Hizo las siguientes apuestas en 5 carreras:

Número de carrera	1	2	3	4	5
Número de apuestas compradas	10	20	10	25	10
Precio por apuesta	$15.40	$16.80	$5.00	$32.50	$3.20

a) ¿Cuál fue el precio promedio por apuesta?
b) Al final de las carreras, recibió las siguientes ganancias por apuesta:

Número de carrera	1	2	3	4	5
Ganancia por apuesta	$1.00	$2.50	$.20	$.80	$.75

¿Cuál fue la ganancia promedio de la inversión en esas apuestas?

3-71 El gerente de operaciones de una fábrica de relojes digitales está pensando abandonar la producción por lotes e introducir una línea continua de montaje. Para poder tomar una decisión acertada, realizó un estudio de tiempos sobre el proceso de lotes. Los datos que reunió presentaron la siguiente distribución de frecuencia del tiempo total de producción de un reloj:

TIEMPO DE PRODUCCION (MIN)	FRECUENCIA	TIEMPO DE PRODUCCION (MIN)	FRECUENCIA
5.00 e inferior	15	7.51– 8.00	58
5.01–5.50	21	8.01– 8.50	63
5.51–6.00	38	8.51– 9.00	27
6.01–6.50	39	9.01– 9.50	21
6.51–7.00	45	9.51–10.00	19
7.01–7.50	51	Mayor a 10.00	14

Determine la mediana. Si el gerente cambia de procesos en cualquier momento, la mediana será superior a 7.75 minutos. ¿Debe cambiar basándose en estos datos?

3-11 AUTOEVALUACION

Las respuestas vienen al final del libro.

V F 1. El valor de las observaciones en el conjunto de datos se tiene en cuenta cuando calculamos su mediana.

V F 2. Cuando la población tiene sesgo negativo o positivo, a menudo es preferible utilizar la mediana como la mejor medida de localización, pues siempre se encuentra entre la media y la moda.

V F 3. Las medidas de tendencia central en un conjunto de datos se refieren al grado de dispersión de las observaciones.

V F 4. Una medida del grado de pico de una curva de distribución es su sesgo.

V F 5. En datos no agrupados, la moda suele emplearse como la medida de tendencia central.

V F 6. Si disponemos de las observaciones en un conjunto de datos por orden descendente, el punto de datos que se halle en la mitad será la mediana del conjunto.

V F 7. Cuando trabajamos con datos agrupados, podemos calcular una media aproximada suponiendo que cada valor en una clase es igual a su marca de clase.

V F 8. El valor que más se repite en un conjunto de datos recibe el nombre de media aritmética.

V F 9. Si la curva de cierta distribución se desvía hacia el extremo izquierdo de la escala de medición sobre el eje horizontal, se dice que la distribución tiene sesgo negativo.

V F 10. Después de agrupar un conjunto de datos en varias clases, podemos identificar la mediana de clase como aquella que posee el mayor número de observaciones.

V F 11. Una media calculada con datos agrupados ofrece siempre una buena estimación del valor verdadero, si bien rara vez es exacta.

V F 12. Podemos calcular una media con cualquier conjunto de datos, una vez que conozcamos su distribución de frecuencia.

V F 13. La moda siempre se encuentra en el punto más alto de una gráfica de una distribución de datos.

V F 14. El número de elementos de una población se denota mediante *n*.

V F 15. Para un arreglo de datos con 50 observaciones, la mediana será el valor de la vigésima quinta observación en el arreglo.

V F 16. Los valores extremos en un conjunto de datos influyen profundamente en la mediana.

V F 17. La diferencia entre las observaciones más grandes y las más pequeñas en un conjunto de datos se llama media geométrica.

18. Si un grupo de datos tiene tan sólo una moda y su valor es menor que el de la media, podremos llegar a la conclusión de que la gráfica de la distribución es:
    a) Simétrica c) Sesgada a la derecha
    b) Sesgada a la izquierda d) Platicúrtica

19. ¿Cuál de las siguientes curvas tendrá el pico más alto?
    a) Leptocúrtica b) Platicúrtica c) Mesocúrtica

20. ¿Cuál es la principal suposición que hacemos al calcular una media a partir de datos agrupados?
    a) Todos los valores son discretos.
    b) Cada valor en una clase es igual a la marca de clase.
    c) Ningún valor ocurre más de una vez.
    d) Cada clase contiene exactamente el mismo número de valores.

21. ¿Cuál de los siguientes enunciados NO es correcto?
    a) Algunos conjuntos de datos no tienen medias.
    b) En los cálculos de la media influyen los valores extremos de datos.
    c) Una media ponderada ha de emplearse cuando es necesario tener en cuenta la importancia de cada valor.
    d) Todos estos enunciados son correctos.

22. ¿Cuál de los siguientes enunciados es el primer paso en el cálculo de la mediana de un conjunto de datos?
    a) Obtener el promedio de los dos valores de la mitad en un conjunto de datos.
    b) Ordenar los datos en un arreglo.
    c) Determinar los pesos relativos de los valores de los datos por orden de importancia.
    d) Ninguno de los anteriores.

23. ¿Cuál de los siguientes enunciados NO es una ventaja del uso de la mediana?
    a) Los valores extremos afectan a la mediana menos intensamente que a la media.
    b) Una mediana puede calcularse para descripciones cualitativas.
    c) La mediana puede calcularse para cualquier conjunto de datos, aun para los que contienen clases abiertas.

d) La mediana es fácil de entender.
e) Todas éstas son ventajas del uso de una mediana.

24. ¿Por qué generalmente es mejor calcular una moda a partir de datos agrupados que a partir de datos no agrupados?
a) Los datos no agrupados tienden a ser bimodales.
b) La moda de datos agrupados será la misma, cualquiera que sea la asimetría de su distribución.
c) Los valores extremos influyen menos en los datos agrupados.
d) Se reduce la probabilidad de que se escoja un valor no representativo.

25. ¿En cuál de estos casos será la moda más útil como indicador de la tendencia central?
a) Cada valor en un conjunto de datos ocurre exactamente una vez.
b) Exceptuados tres, todos los valores ocurren una vez en el conjunto de datos; tres ocurren 100 veces cada uno.
c) Todos los valores en un conjunto de datos ocurren 100 veces cada uno.
d) Cada observación en un conjunto de datos tiene el mismo valor.

26. ¿Cuál de los siguientes es un ejemplo de parámetro?
a) $\bar{x}$ c) μ e) b y c pero no a
b) n d) Todos los anteriores

27. ¿Cuál de las siguientes designaciones y enunciados NO es una medida de tendencia central?
a) Media geométrica c) Moda e) Todas éstas son medidas de la tendencia central
b) Mediana d) Media aritmética

28. Cuando una distribución es simétrica y tiene una moda, el punto más alto en la curva se llama:
a) Intervalo c) Mediana e) Todo lo anterior
b) Moda d) Media f) b, c y d, pero no a

29. Cuando nos referimos a una curva que se desvía hacia el extremo izquierdo, deberemos llamarla:
a) Simétrica c) De sesgo positivo e) Nada de lo anterior
b) Sesgada a la derecha d) Todo lo anterior

30. Si una curva puede dividirse en partes iguales que son imágenes de espejo, será ________________. Si no puede ser dividida en esa forma, será ________________.

31. El símbolo $\bar{x}$ denota la media de ________________. μ denota la media de ________________.

32. Asignar enteros consecutivos de valor pequeño a las marcas de clase durante el cálculo de la media se llama ________________.

33. Cuando trabajamos con cantidades que cambian en un periodo de tiempo, es mejor calcular una ________________ media que una ________________ media.

34. Si dos valores en un grupo de datos ocurren más a menudo que otros cualesquiera, la distribución de los datos será ________________.

35. El grado en que los valores de una distribución se agrupan es una medida de ________________.

3-12 CASO CONCEPTUAL
(Northern White Metals Company)

A Dick le impresionó muchísimo la cantidad de información que, con ayuda de dos renuentes archivistas, había entresacado de gavetas repletas. Las gavetas servían de depósito para documentos que eran demasiado importantes y no podían desecharse, pero carecían de utilidad inmediata. Sin duda, las actuales facturas, pedidos y demás material pertinente debían estar siempre a la mano; pero Dick sonrió pensando si alguna vez se necesitaría un acuse de recibo de un pedido con fecha de 12 de mayo de 1954.

El primer paso había consistido en ordenar y actualizar el sistema de registros: todos los archivos que tenían una antigüedad de más de tres años fueron llevados al sótano donde se conservarían en forma permanente. El material carente de importancia fue eliminado por completo. Una vez que disminuyó el material en desorden, Dick pudo organizar la información necesaria en una forma que la parecía idónea para la presentación en las oficinas centrales de Segue. Los registros fueron revisados, y Dick efectuó un análisis minucioso del volumen de ventas, del volumen de producción, así como de las fluctuaciones en el precio de las materias primas y los patrones de compra.

Lleno de ansiedad, Dick llevó los datos que había acumulado y después los organizó con el jefe de dibujo, George Barbour, quien sentía una gran afición por el arte y una notable habilidad para el dibujo mecánico. Se elaboró una impresionante serie de tablas y gráficas, y Dick pensó que los detalles favorecerían una presentación mucho más eficaz. Pero George no compartía esa opinión.

"Dick, según tengo entendido, has hecho un excelente trabajo al juzgar el estado de la compañía basándote en las cifras de ventas y producción. Pero en cuanto a tu presentación pienso que tenemos un verdadero problema", comentó con cautela.

"¿Por qué crees eso, George?", preguntó Dick, aunque empezaba a entender la razón mientras examinaba cuidadosamente un montón de información.

"Todo esto es demasiado complejo y detallado. No hay ningún dato que sobresalga y nos diga en forma concisa: 'Esta es la situación' ", prosiguió George. "Tengo la impresión de que, para cuando llegues a la parte central de tu exposición, todos estarán muy confundidos. Este es un magnífico material de apoyo, pero a mi juicio debería haber alguna medida resumida de esas áreas que realmente describa la situación."

Dick asintió, pues ya conocía muy bien el interés de George y su sagacidad. "Pienso que en esto tienes toda la razón", replicó con un tono de gratitud. "Voy a repasar todo esto y veré qué puedo hacer para mejorar la presentación."

Dick ha identificado tres áreas principales (volumen de ventas, volumen de producción y compras de materias primas) sobre las cuales aportará algunos datos y analizará las tendencias de los tres últimos años. Ello le permitirá hacerse una idea más clara de la situación actual de la compañía y de cómo se llegó a ella. George sugirió que algunas medidas resumidas de la abundante información recopilada por Dick podrían hacer más interesante y significativa su presentación. ¿Qué puede hacer para describir más claramente las ventas, las ventas por clase de cliente, las ventas por región, el volumen de producción, el análisis de departamentos, etc., de manera que la administración corporativa esté mejor preparada para abordar el problema de describir gráficamente la dirección de la empresa en un futuro próximo?

3-13 EJERCICIO CON LA BASE DE DATOS

Ante el hecho de la disminución de las ventas, Fred Walker preparó una estrategia general de mercadotecnia para los trineos de madera luego de intercambiar opiniones con su fuerza de ventas. Su subordinado, Dan Riffen, empezaba a sentir los efectos secundarios de la nueva competencia de mercado en Cold River.

"Esos jóvenes imberbes de la computación creen que pueden decirme cómo dirigir la fuerza de ventas", exclamó Fred. "¿Pero qué es lo que saben de empujar un trineo? Seguramente ni siquiera saber encerar un par de esquís y mucho menos qué parte del trineo Rough Rider es la del frente."

"Ampliemos esa sección en el plan de mercadotecnia que se refiere a las promociones en las pequeñas tiendas de juguetes. Agreguemos al paquete de cortesía del minorista un juego gratis de Cold River. Y con eso bastará", exclamó Fred con tono ufano.

Fred había diseñado una estrategia de mercadotecnia basándose en dos suposiciones:

1. La parte medular de las ventas de Cold River es la tienda pequeña.
2. La fuerza de ventas es la principal ventaja de la empresa.

Fred dirigió su campaña a los vendedores y a las tiendas de barrio donde las ventas estaban mermando. Planeó el ataque en dos frentes.

"Primero", sugirió, "debemos motivar a la fuerza de ventas. Los convocaremos a las oficinas centrales para una reunión por la noche en el teatro de la ópera en Central City. Ofreceremos manjares de lo más exquisito y ordenaremos que un grupo de lindas muchachas hagan una demostración de los trineos. Luego asestaremos el golpe definitivo con nuestra magnífica promoción de ventas y ellos no tendrán más remedio que salir a la calle a vender trineos."

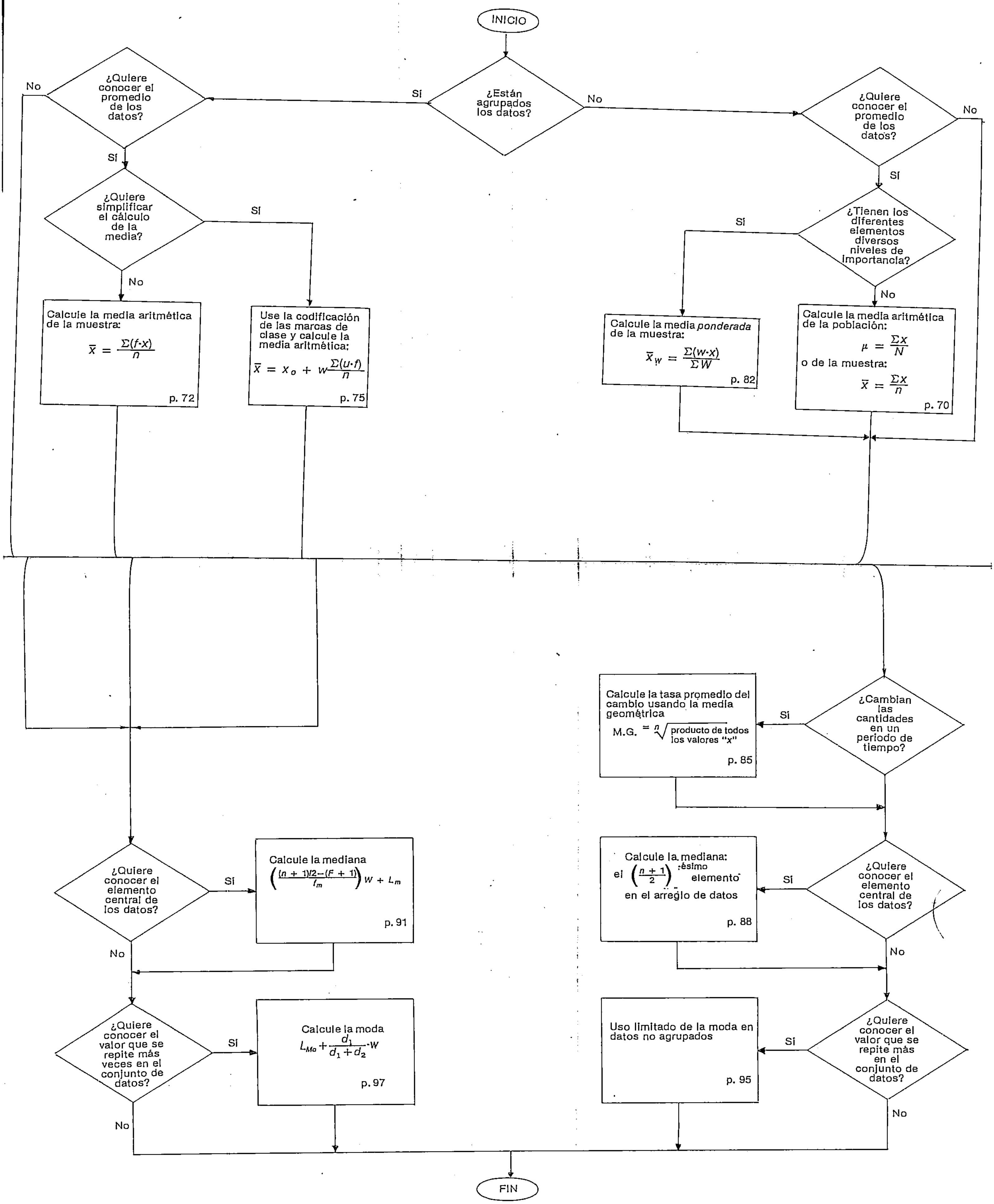
"El siguiente paso", explicó Fred, "consiste en convencer a los dueños de las tiendas. Comenzaremos en grande con un nuevo calendario de deportes invernales y un descuento a pedidos mayores de 30 unidades. Si con esto no mejoran las ventas, ofreceremos una semana gratis en Aspen, con todos los gastos pagados, al dueño de la tienda que logre las ventas más altas."

"Esos muchachos no tienen la menor oportunidad con mi enorme experiencia. Les prometo que para Navidad ya no estarán aquí."

PROBLEMAS Y PREGUNTAS

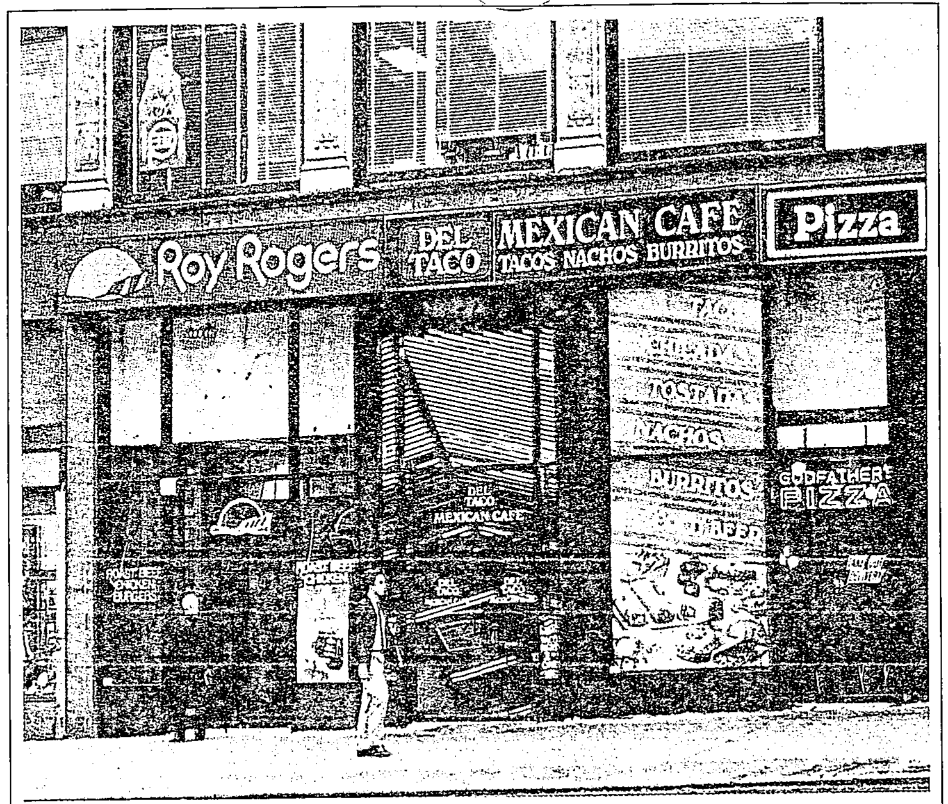
1. Calcule la media, la mediana y la moda de los datos referentes al tamaño de los pedidos anuales, que se incluyeron en el capítulo 2.
2. En los últimos tres años han ido disminuyendo las ventas. ¿Qué ha ocurrido al tamaño medio de las órdenes? ¿Son compatibles entre sí esas tendencias? Explique su respuesta.
3. Si las tiendas pequeñas hacen pedidos pequeños y las tiendas grandes colocan grandes pedidos, ¿qué sucede con la importancia relativa de las tiendas pequeñas? ¿Está obrando correctamente Fred al promover las ventas exclusivamente en los locales pequeños?

3-14 DIAGRAMA DE FLUJO



4 Medición de la variabilidad

OBJETIVOS: En este capítulo ponemos fin al estudio de la estadística descriptiva examinando los métodos que nos permiten medir la tendencia de un grupo de datos a esparcirse o diseminarse. Supongamos que un nuevo sistema de frenado detiene el vehículo de prueba a una distancia promedio de 30 m, distancia aceptable según los requisitos de seguridad en carretera. Pero ahora supongamos que en los datos de la prueba encontramos muchas lecturas de distancia de frenado de 80 m aproximadamente, distancia totalmente inaceptable. En este caso, el uso del promedio y el hecho de prescindir de la tendencia de los datos a esparcirse resultarán muy peligrosos. En el presente capítulo se explicarán métodos para manejar esta variabilidad.



El vicepresidente de mercadotecnia de una cadena de locales de comida de preparación rápida está estudiando las ventas de 100 locales situados en el distrito oriental y ha preparado la siguiente distribución de frecuencia de las ventas anuales:

VENTAS (en miles)	FRECUENCIA	VENTAS (en miles)	FRECUENCIA
700– 799	4	1,300–1,399	13
800– 899	7	1,400–1,499	10
900– 999	8	1,500–1,599	9
1,000–1,099	10	1,600–1,699	7
1,100–1,199	12	1,700–1,799	2
1,200–1,299	17	1,800–1,899	1

Al vicepresidente le gustaría comparar el distrito oriental con los otros tres del país. Para hacerlo sumarizará la distribución, pero procurando al mismo tiempo recabar más información que la simple medida de tendencia central. Este capítulo explica cómo ese ejecutivo puede medir la variabilidad en una distribución y hacerse así una mejor idea de los datos.

4-1 MEDIDAS DE DISPERSION

Necesidad de medir la dispersión o variabilidad

En el capítulo 3, dijimos que dos conjuntos de datos pueden tener la misma localización central y, no obstante, ser muy distintos si uno se halla más disperso que el otro. Ello se aplica a las tres distribuciones de la figura 4-1. La media de las tres curvas es la misma, pero la curva A tiene menor dispersión (o *variabilidad*) que la curva B, y ésta a su vez presenta menor variabilidad que la curva C. Si medimos sólo la media de las tres distribuciones, no captaremos una importante diferencia que hay entre las tres curvas. Como en cualquier tipo de datos, la media, la mediana y la moda nos indican sólo parte de lo que necesitamos conocer en torno a las características de los datos. Para mejorar nuestro conocimiento del patrón de datos, es preciso que midamos además su *dispersión*, o sea su *variabilidad*.

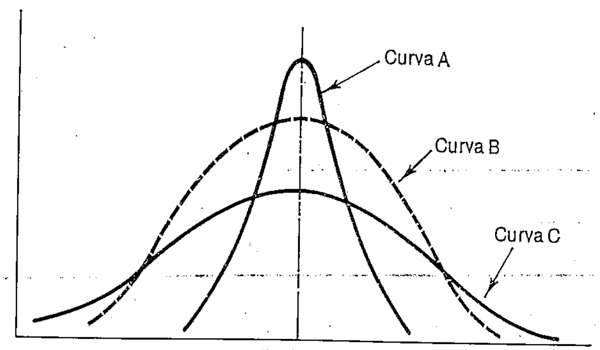


FIGURA 4-1 Tres curvas con la misma media pero con variabilidades diferentes

Usos de las medidas de dispersión

¿Por qué la dispersión de la distribución es una característica tan importante de entender y medir? Primero, nos suministra información complementaria que nos permite juzgar la confiabilidad de nuestra medida de tendencia central. Si los datos están ampliamente dispersos, como los de la curva C en la figura 4-1, la localización central será menos representativa de los datos en su conjunto de lo que sería en el caso de datos que se acumulasen más alrededor de la media, como sucede en la curva A. Segundo, puesto que se trata de problemas típicos de datos sumamente dispersos, se requiere la capacidad de reconocer que los datos están muy dispersos pues de lo contrario no podremos abordar esos problemas. Y, tercero, tal vez queramos comparar las dispersiones de varias muestras. Si no conviene tener una amplia dispersión de valores respecto al centro o si esa dispersión entraña un riesgo inaceptable, deberemos ser capaces de reconocerlo y no escoger las distribuciones que presentan la máxima dispersión.

Aplicaciones financieras

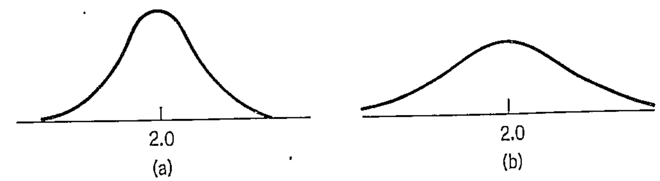
A los analistas financieros les interesa la dispersión de las ganancias de una empresa. Las utilidades con una fuerte dispersión, o sea las que incluyen desde niveles extremadamente altos hasta los muy bajos e incluso negativos, indican un riesgo mayor para los accionistas y acreedores que las utilidades que permanecen relativamente estables. De manera análoga, los expertos en control de calidad

Uso del control de calidad

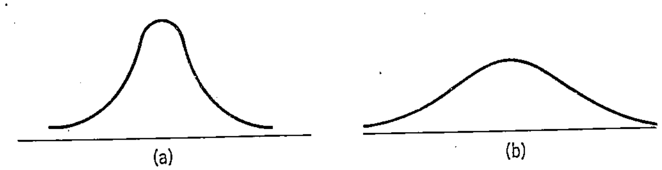
analizan la dispersión de los niveles de calidad de un producto. Un medicamento que tiene una pureza promedio pero que fluctúa entre muy alto grado de pureza e impureza puede constituir un peligro para la vida del paciente.

EJERCICIOS

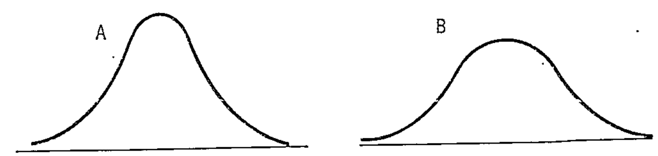
4-1 Una compañía que aplica 2 métodos diferentes para surtir los pedidos de sus clientes encontró las siguientes distribuciones del tiempo de entrega en ambos métodos, basándose en registros anteriores. ¿Cuáles métodos recomendaría usted a partir de la evidencia disponible?



4-2 ¿Para cuál de las siguientes distribuciones es la media más representativa de los datos en su conjunto? ¿Por qué?



4-3 Para medir el aprovechamiento escolar los educadores necesitan conocer, mediante pruebas, los niveles de conocimiento y la capacidad de los alumnos. Pueden planear mejor los programas de estudio si tienen en cuenta las diferencias individuales de los alumnos. Las curvas siguientes representan distribuciones basadas en las calificaciones promedio logradas en 2 pruebas diferentes. ¿Cuál es, a su juicio, la mejor para los educadores que van a elaborar el plan de estudios?



4-4 ¿Cuál de los siguientes enunciados no es el motivo por el cual se mide la dispersión de una distribución?

a) Ofrece una indicación de la confiabilidad de la medida de tendencia central.
b) Nos permite comparar varias muestras que tienen promedios similares.
c) Usa más datos al describir una distribución.

d) Centra la atención sobre los problemas que entraña una variabilidad muy grande o muy pequeña en las distribuciones.

4-5 Entre las 3 curvas de la figura 4-1, escoja la que describa mejor la distribución de valores para las edades de los siguientes grupos: miembros del Congreso, miembros recién electos de la Cámara de diputados, el presidente de los principales comités del Congreso. Al hacer su selección, prescinda de la media común de las curvas de la figura 4-1 y tenga en cuenta tan sólo la variabilidad de las distribuciones. Exponga brevemente los motivos de su elección.

4-6 ¿Cómo piensa que el concepto de variabilidad podría aplicarse a una investigación realizada por la Federal Trade Commission (FTC) respecto a la fijación de precios por parte de un grupo de fabricantes?

4-7 Diga cuál de las 3 curvas de la figura 4-1 describe mejor la distribución de las siguientes características de diversos grupos. Haga su elección basándose exclusivamente en la variabilidad de las distribuciones. Explique brevemente una razón de cada elección.

a) El número de puntos anotados por cada jugador de un equipo profesional de baloncesto durante una temporada de 80 juegos.
b) El sueldo de 100 personas que trabajan en empleos más o menos equivalentes del gobierno federal.
c) El promedio de calificaciones de 15,000 alumnos de una de las principales universidades estatales.
d) El sueldo de 100 personas que laboran en trabajos más o menos equivalentes de una corporación privada.
e) La calificación promedio de los alumnos de una de las principales universidades estatales que han sido aceptados en una escuela de posgrado.
f) El porcentaje de lanzamientos hechos por los jugadores de una liga profesional de baloncesto durante una temporada de 80 juegos.

4-2 DISPERSION: medidas de distancia

Tres medidas de distancia

La dispersión puede medirse en términos de la diferencia existente entre dos valores seleccionados del conjunto de datos. En la presente sección, estudiaremos tres de las llamadas *medidas de distancia*: el intervalo, el intervalo de interfractil y la desviación de cuartil.

Intervalo

Definición y cálculo del intervalo

Como dijimos en el capítulo 2, el intervalo es la *diferencia entre el valor más alto y el más bajo observados*. En forma de ecuación, esto podemos expresarlo así:

$$\text{Intervalo} = \begin{array}{c}\text{Valor de la observación}\\ \text{más alta}\end{array} - \begin{array}{c}\text{Valor de la observación}\\ \text{más baja}\end{array} \qquad [4\text{-}1]$$

Aplicando la ecuación anterior, comparamos los intervalos de los pagos anuales que Blue Cross-Blue Shield hacen a los dos hospitales incluidos en la tabla 4-1.

El intervalo de los pagos anuales a Cumberland es de \$1,883,000 – \$863,000 = \$1,020,000. Para Valley Falls, el intervalo es de \$690,000 – \$490,000 = \$200,000.

TABLA 4-1 Pagos anuales de Blue Cross-Blue Shield (en miles)

Cumberland	863	903	957	1,041	1,138	1,204	1,354	1,624	1,698	1,745	1,802	1,883
Valley Falls	490	540	560	570	590	600	610	620	630	660	670	690

Características del intervalo

El intervalo es fácil de entender y calcular, pero es escasa su utilidad como medida de dispersión. El intervalo incluye únicamente los valores máximo y mínimo de una distribución, sin tener en cuenta ninguna otra observación dentro del conjunto de datos. De ahí que ignore la naturaleza de la variación entre todas las demás observaciones, siendo afectado profundamente por los valores extremos. Mide sólo dos valores, por lo cual el intervalo tiende a cambiar drásticamente entre una muestra y la siguiente en una población dada, pese a que los valores que caen entre el valor más alto y el más bajo pueden ser muy similares. No se olvide que las distribuciones abiertas no tienen intervalo, dado que en la clase abierta no existe valor "máximo" ni "mínimo".

Intervalo de interfractil

Fractiles

En una distribución de frecuencia, una fracción o proporción de los datos se encuentra en un *fractil* o debajo de él. Así, la mediana es el .5 fractil porque la mitad del conjunto de datos son menores o iguales que este valor. El lector advertirá que los fractiles se parecen a los porcentajes. En toda distribución, 25% de los datos se halla en el .25 fractil o debajo de él; de manera similar, 25% de los datos se sitúan en el vigésimo quinto percentil o debajo de él. El *intervalo de interfractil* es una medida de la dispersión entre los valores del fractil.

Significado del intervalo del interfractil

Cálculo del intervalo del interfractil

Supóngase que deseamos obtener el intervalo de interfractil entre el primero y segundo *tercios* de lo que Cumberland recibe de Blue Cross-Blue Shield. Empezamos dividiendo la observación en tercios, como hicimos en la figura 4-2. Cada tercio contiene cuatro elementos (1/3 del total de doce elementos). Por tanto, 33 1/3 de los elementos se encuentra en \$1,041,000 o debajo de esta cifra, y 66 2/3% son iguales o menores que \$1,624,000. Ahora ya podemos calcular el intervalo del interfractil entre 1/3 y 2/3 fractiles al restar el valor \$1,041,000 al valor \$1,624,000. Esta diferencia de \$583,000 es la dispersión entre la parte superior del primer tercio de pagos y la parte superior del segundo tercio.

TABLA 4-2 Pagos anuales de Blue Cross-Blue Shield al Comberland Hospital (en miles)

PRIMER TERCIO	SEGUNDO TERCIO	TERCER TERCIO
863	1,138	1,698
903	1,204	1,745
957	1,354	1,802
1,041 ← 1/3 de fractil	1,624 ← 2/3 de fractil	1,883

Fractiles especiales: deciles, cuartiles y percentiles

Los fractiles pueden tener nombres especiales según el número de partes iguales en que dividen los datos. Pueden dividirse en diez partes iguales denominadas *deciles*. Los *cuartiles* dividen los datos en cuatro partes. Los *percentiles* dividen los datos en 100 partes iguales. Probablemente el lector se haya encontrado con percentiles en las calificaciones de sus pruebas. Sabe que obtuvo una puntuación en el septuagésimo quinto percentil, 3/4, o sea 75%, de todos los que hicieron la prueba no lograron calificaciones más altas.

Intervalo de intercuartil y desviación de cuartil

Cálculo del intervalo del intercuartil

El intervalo de intercuartil mide aproximadamente la distancia de la mediana que debemos recorrer en ambos lados antes de poder incluir una mitad de los valores del conjunto de datos. Para calcular este intervalo, dividimos los datos en cuatro partes, cada una de las cuales contiene 25% de los elementos de la distribución. Los *cuartiles* son, pues, los valores máximos en las cuatro partes, y el *intervalo del intercuartil* es la diferencia entre los valores del primer y tercer cuartiles:

$$\text{Intervalo de intercuartil} = Q_3 - Q_1 \qquad [4\text{-}2]$$

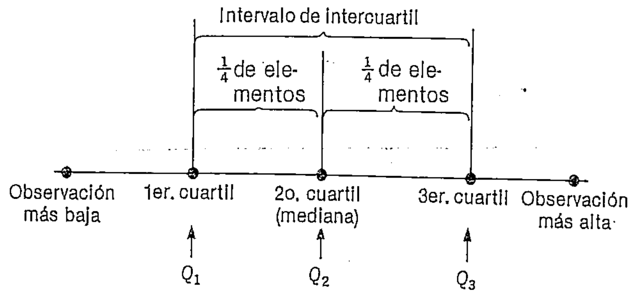


FIGURA 4-2 Intervalo de intercuartil

La figura 4-2 muestra el concepto del intervalo del intercuartil en forma gráfica. Nótese en esa figura que la amplitud de cada uno de los cuatro cuartiles *no necesariamente es idéntica*.

En la figura 4-3, otro ejemplo de cuartiles, éstos dividen el área bajo la distribución en cuatro partes iguales, cada una de las cuales contiene 25% del área.

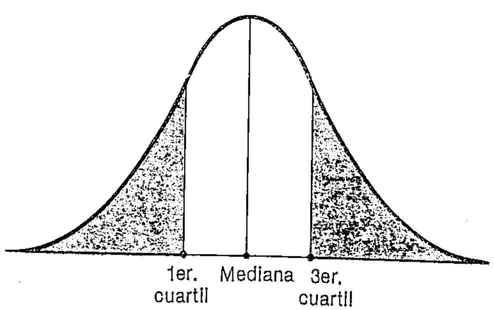


FIGURA 4-3 Cuartiles

Una mitad intervalo de intercuartil es una medida denominada *desviación del cuartil*:

$$\text{Desviación del cuartil} = \frac{Q_3 - Q_1}{2} \qquad [4\text{-}3]$$

Desviación del cuartil

La desviación del cuartil mide, pues, el intervalo *promedio* de un cuarto de los datos. Es representativa de todos ellos, ya que se obtiene tomando un promedio de la mitad intermedia de los datos en vez de escoger una de las cuartas partes.

Problema que ejemplifica el intervalo del intercuartil y la desviación del cuartil

Encontremos el intervalo del intercuartil y la desviación del cuartil de los pagos anuales que Blue Cross-Blue Shield hace anualmente a Cumberland y que aparecen en la tabla 4-1. Comenzamos dividiendo los elementos en cuatro partes iguales, como hicimos en la tabla 4-3. En ésta vemos que el tercer cuartil es $1,698,000 y que el primer cuartil es $957,000. Aplicando la ecuación 4-2 obtenemos el intervalo del intercuartil, el cual es $741,000:

$$\begin{aligned}\text{Intervalo de intercuartil} &= Q_3 - Q_1 \qquad [4\text{-}3]\\ &= 1{,}698 - 957\\ &= 741 \text{ mil dólares}\end{aligned}$$

y la *desviación del cuartil* es $370,500:

$$\begin{aligned}\text{Desviación de cuartil} &= \frac{Q_3 - Q_1}{2} \qquad [4\text{-}2]\\ &= \frac{1{,}698 - 957}{2}\\ &= \frac{741}{2}\\ &= 370.5 \text{ mil dólares}\end{aligned}$$

TABLA 4-3 Pagos anuales de Blue Cross-Blue Shield al Cumberland Hospital (en miles)

PRIMER CUARTO	SEGUNDO CUARTO	TERCER CUARTO	ULTIMO CUARTO
863	1,041	1,354	1,745
903	1,138	1,624	1,802
957 ← primer cuartil	1,204	1,698 ← tercer cuartil	1,883

Ventajas del intervalo del intercuartil y la desviación del cuartil

Igual que el intervalo, el intervalo de intercuartil y la desviación del cuartil se basan en dos únicos valores del conjunto de datos. Aunque son más complicados de calcular que el intervalo, evitan los valores extremos utilizando únicamente la mitad intermedia de los datos. De ahí su ventaja neta sobre el intervalo, el cual se ve afectado por los valores extremos.

EJERCICIOS

4-8 A continuación se transcriben los precios totales (en dólares) cobrados el martes por 20 taxis que pertenecen a la empresa City Transit, Ltd.:

98	120	80	132	126	144	92	90	124	140
90	130	52	122	129	190	89	112	123	148

Calcule el intervalo de los datos anteriores y diga si, en su opinión, conviene usarlo como una medida útil de dispersión en este caso.

4-9 He aquí las calificaciones de los alumnos en un examen de historia. Calcule la calificación que representa el octagésimo percentil:

97	89	87	84	75	52	100	98	95	82
91	94	88	86	73	68	80	85	76	79

4-10 Para los siguientes datos, calcule:

a) El intervalo del intercuartil
b) La desviación del cuartil

97	72	87	57	39	81	70	84	93	79
84	81	65	97	75	72	84	46	94	77

4-11 Para la siguiente muestra calcule:

a) El intervalo
b) El intervalo del interfractil entre el vigésimo y el octagésimo percentil
c) El intervalo del intercuartil
d) La desviación del cuartil
e) El intervalo del interfractil entre el primero y el segundo cuartiles

Compare las partes (d) y (e).

2,696	2,880	2,575	2,748	2,762	2,572	3,233	2,733	2,890	2,878
3,100	3,321	2,693	2,865	2,784	3,296	2,977	2,090	2,905	3,350

4-12 Una compañía de transportes lleva el siguiente registro del tiempo (redondeado a la centésima de minuto más cercana) que sus camiones esperaron para ser descargados. Calcule la desviación del cuartil y el intervalo del intercuartil para sus datos.

.10	.20	.38	.45	.50	.61	.71	.83	.88	.98	1.02	1.18
.12	.28	.40	.46	.53	.68	.73	.84	.91	1.00	1.10	1.20
.15	.32	.42	.49	.59	.70	.75	.86	.96	1.01	1.15	1.24

4-13 Una fábrica de aparatos domésticos ha desarrollado una nueva combinación de mezcladora y olla. En una demostración de mercadotecnia, una encuesta de precios reveló que la mayor parte de las personas muestreadas estaban dispuestas a pagar unos $60, con una desviación de cuartil extraordinariamente pequeña: $7.20. Con el propósito de reproducir los resultados, se repitieron la demostración y la encuesta concomitante. El departamento de mercadotecnia confiaba obtener una desviación del cuartil todavía menor. Los datos se transcriben en seguida. ¿Se cumplieron sus expectativas?

PRECIO (EN DOLARES)

52	35	48	46	43	40	61	49	57	58	65	46
72	69	38	37	55	52	50	31	41	60	45	41

4-14 Una compañía de seguros está examinando la conveniencia de comprar una nueva flotilla de automóviles. El director del departamento de finanzas, muestreó a 40 empleados para determinar el número de millas que cada auto recorrió en un periodo de 1 año. Los resultados del estudio se incluyen a continuación. Calcule la desviación del cuartil y el intervalo del intercuartil.

3,600	4,200	4,700	4,900	5,300	5,700	6,700	7,300
7,700	8,100	8,300	8,400	8,700	8,700	8,900	9,300
9,500	9,500	9,700	10,000	10,300	10,500	10,700	10,800
11,000	11,300	11,300	11,800	12,100	12,700	12,900	13,100
13,500	13,800	14,600	14,900	16,300	17,200	18,500	20,300

4-15 El New Mexico State Highway Department tiene la obligación de mantener en buenas condiciones todas las carreteras del estado. Una medida de ellas es el número de grietas que hay en cada 30 m de la carretera. Con el estudio anual del departamento se construyó la siguiente distribución:

GRIETAS POR CADA 100 PIES

2	5	6	7	7	8	9	10	10	11
12	12	12	13	13	13	14	14	15	15
15	15	16	16	17	17	18	19	19	20

Calcule el intervalo de interfractil entre el vigésimo, cuadragésimo y octagésimo percentiles.

4-16 Teodoro Rentería es un analista estadístico que "reporta" directamente a los niveles más altos de la gerencia en la corporación. Ayudó a diseñar el eslogan de la compañía: "Si no puede encontrar la solución, se la damos nosotros". El señor Rentería acaba de recibir algunos datos que le inquietan mucho: el volumen mensual monetario de contratos de investigación ganados por la compañía el año anterior. En teoría, esas cifras mensuales deberían ser bastante estables, dado que una excesiva fluctuación en el trabajo por realizar puede originar desorden en la contratación de personal y despido de empleados. A continuación se dan los datos de Ted (en miles de dólares):

253	104	633	57	500	201
43	380	467	162	220	302

Calcule lo siguiente:

a) El intervalo del interfractil entre el segundo y el octavo deciles
b) La mediana, Q_1 y Q_3
c) La desviación del cuartil

4-3 DISPERSION: medidas de desviación promedio

Dos medidas de la desviación promedio

Las descripciones más completas de la dispersión son aquellas que se refieren a la desviación promedio respecto a alguna medida de tendencia central. Dos de esas medidas son importantes en nuestro estudio de la estadística: la *variancia* y la *desviación estándar*. Ambas nos indican una distancia promedio de cualquier observación en el conjunto de datos a partir de la media de la distribución.

Desviación absoluta promedio

Significado de la desviación absoluta promedio

Nos haremos una mejor idea de la variancia y de la desviación estándar si nos concentramos primero en lo que los estadísticos llaman *desviación absoluta promedio*. Para calcularla comenzamos obteniendo la media de la muestra. Luego determinamos el valor absoluto de la diferencia entre cada elemento del conjunto de datos y la media. En otras palabras, restamos la media a cada valor del conjunto e ignoramos el signo (positivo o negativo), suponiendo con ello que todos sean positivos. Por último, sumamos todas las diferencias y las dividimos entre el número total de elementos de la muestra.

En su expresión simbólica, la fórmula con que se calcula la desviación absoluta promedio tiene esta forma:

$$\text{Desviación absoluta promedio} = \frac{\Sigma|x-\mu|}{N} \text{ para una población} \qquad [4\text{-}4]$$

y parecida a ésta

$$\text{Desviación absoluta promedio} = \frac{\Sigma|x-\bar{x}|}{n} \text{ para una muestra} \qquad [4\text{-}5]$$

donde:

- x = el elemento o la observación
- μ = la media de la población
- N = número de elementos de la población
- $\bar{x}$ = media muestral
- n = número de elementos de la muestra

Recuérdese que Σ significa "la suma de todos los valores". En este caso, pueden ser $|x-\mu|$ o bien $|x-\bar{x}|$. Adviértanse también las líneas rectas que rodean $|x-\mu|$ y $|x-\bar{x}|$, que indican que queremos el *valor absoluto* de esa distancia (expresada en números positivos, no negativos). Ello significa que, si la distancia $x-\bar{x}$ es -10, entonces el valor absoluto será 10. Y el valor absoluto de -25 es 25.

Cálculo de la desviación absoluta promedio

Calculemos ahora la desviación absoluta promedio de los pagos anuales de Blue Cross-Blue Shield a Cumberland en la tabla 4-1. Ante todo calculamos la media:

$$\bar{x} = \frac{\Sigma x}{n} \qquad [3\text{-}2]$$

$$= \frac{863+903+957+1{,}041+1{,}138+1{,}204+1{,}354+1{,}624+1{,}698+1{,}745+1{,}802+1{,}883}{12}$$

$$= \frac{16{,}212}{12}$$

$$= 1{,}351 \text{ mil dólares}$$

Aplicando el proceso gradual descrito en la tabla 4-4, obtenemos la desviación absoluta de todas las observaciones partiendo de esta media de \$1,351,000. A continuación dividimos la suma de las desviaciones absolutas entre el número de elementos de la muestra para obtener el valor de la desviación absoluta promedio:

$$\text{Desviación absoluta promedio} = \frac{\Sigma|x-\bar{x}|}{n} \qquad [4\text{-}5]$$

$$= \frac{4{,}000}{12}$$

$$= 333.3 \text{ mil dólares}$$

Características de la desviación absoluta promedio

Esta desviación absoluta promedio es una mejor medida de la dispersión que los intervalos que ya hemos calculado, pues tiene en cuenta todas las observaciones. Pondera por igual cada elemento e indica a qué distancia de la media se halla en promedio cada observación. Pese a esa ventaja, razones técnicas que exceden el ámbito de esta obra hacen que se use poco este método de desviación promedio.

TABLA 4-4 Determinación de la desviación absoluta promedio de los pagos de Blue Cross-Blue Shield a Cumberland Hospital (en miles)

OBSERVACION (x) (1)		MEDIA ($\bar{x}$) (2)		DESVIACION ($x-\bar{x}$) (1) − (2)	DESVIACION ABSOLUTA ($\|x-\bar{x}\|$) |(1) − (2)|
863	−	1,351	=	−488	488
903	−	1,351	=	−448	448
957	−	1,351	=	−394	394
1,041	−	1,351	=	−310	310
1,138	−	1,351	=	−213	213
1,204	−	1,351	=	−147	147
1,354	−	1,351	=	3	3
1,624	−	1,351	=	273	273
1,698	−	1,351	=	347	347
1,745	−	1,351	=	394	394
1,802	−	1,351	=	451	451
1,883	−	1,351	=	532	532
16,212 ← Σx					4,000 ← $\Sigma\|x-\bar{x}\|$

Variancia de la población

Variancia

En las dos siguientes secciones, vamos a centrarnos en la población más que en las muestras. Comenzaremos con el hecho de que cada población tiene una variancia, la cual se representa mediante σ^2 (*sigma al cuadrado*).

Relación de la variancia con la desviación absoluta promedio

La variancia de la población se parece a una desviación absoluta promedio calculada para una población entera. Sólo que en este caso estamos utilizando la suma de los *cuadrados* de las distancias entre la media y cada elemento, divididos entre el número total de los elementos que hay en la población. Al elevar al cuadrado cada distancia, automáticamente hacemos positivos todos los números y, por consiguiente, no necesitamos tener en cuenta el valor absoluto de cada desviación.

Fórmula de la variancia

La fórmula con que se calcula la variancia es similar a la ecuación 4-4. Pero esta vez, como estamos obteniendo el cuadrado de la distancia promedio entre la media y cada elemento de la población, elevamos al cuadrado las diferencias de $x - \mu$:

$$\sigma^2 = \frac{\Sigma(x-\mu)^2}{N} = \frac{\Sigma x^2}{N} - \mu^2 \qquad [4\text{-}6]$$

donde:

- σ^2 = la variancia de la población
- x = el elemento u observación
- μ = media de la población
- N = número total de elementos de la población
- Σ = suma de todos los valores $(x-\mu)^2$ o sea todos los valores x^2.

En la ecuación 4-6, la expresión de la mitad, $\frac{\Sigma(x-\mu)^2}{N}$, es la definición de σ^2. La última expresión, $\frac{\Sigma x^2}{N} - \mu^2$ es, *matemáticamente*, equivalente a la definición, pero a menudo conviene más utilizarla si debemos obtener el valor de σ^2, puesto que entonces no tenemos que calcular la desviación de la media. No obstante, cuando los valores de x son grandes y los valores $x - \mu$ son pequeños, quizá sea preferible usar la expresión de la mitad, $\frac{\Sigma(x-\mu)^2}{N}$, para calcular σ^2. Antes de emplear esta fórmula en un ejemplo, hemos de explicar un problema muy importante referente a la variancia. Al resolverlo aprenderemos lo que es la desviación estándar y cómo calcularla. Después podemos ocuparnos de la variancia propiamente dicha.

Las unidades en que se expresa la variancia causan problema

En páginas anteriores, cuando calculamos el intervalo, la desviación del cuartil y la desviación absoluta promedio, las respuestas fueron expresadas en las mismas unidades que los datos. (En nuestros ejemplos, las unidades fueron "miles de dólares de pagos".) Sin embargo, en el caso de la variancia, las unidades fueron los *cuadrados de las unidades* de los datos; por ejemplo, "dólares al cuadrado" o "cuadrados de dólares". Estas dos últimas expresiones no son intuitivamente claras ni fáciles de interpretar. Por tal razón, hemos de introducir un cambio importante en la variancia para obtener una medida útil de la desviación, la cual no nos plantee un problema con las unidades de medida y que, por lo mismo, resulte menos confusa. **Esta medida recibe el nombre de desviación estándar y es la raíz cuadrada de la variancia.** La raíz cuadrada de \$100 al cuadrado es \$10, porque tomamos la raíz cuadrada del valor y de las unidades en que se mide. Así pues, la desviación estándar se da en unidades que son las mismas que los datos originales.

Relación de la desviación estándar con la variancia

Desviación estándar de la población

La desviación estándar de la población, o σ, es simplemente la raíz cuadrada de la variancia de la población. La variancia es el promedio de los cuadrados de las distancias de las observaciones hechas a partir de la media, **por lo cual la desviación estándar es la raíz cuadrada del promedio de los cuadrados de las distancias de las observaciones de la media.** La variancia se expresa en el cuadrado de las unidades usadas en los datos; en cambio, la desviación estándar está en las mismas unidades que las empleadas en los datos. He aquí la fórmula de la desviación estándar:

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\Sigma(x-\mu)^2}{N}} = \sqrt{\frac{\Sigma x^2}{N} - \mu^2} \qquad [4\text{-}7]$$

donde

- x = la observación
- μ = la media de la población
- N = el número total de elementos en la población
- Σ = la suma de todos los valores $(x-\mu)^2$, o sea todos los valores x^2
- σ = la desviación estándar de la población
- σ^2 = la variancia de la población

Empleo de la raíz cuadrada positiva

La raíz cuadrada de un número positivo puede ser positiva o negativa, puesto que $a^2 = (-a)^2$. Cuando tomamos la raíz cuadrada de la variancia para calcular la desviación estándar, los estadísticos no tienen en cuenta más que la raíz cuadrada positiva.

Cálculo de la desviación estándar

Para calcular la variancia o la desviación estándar, construimos una tabla utilizando todos los elementos de la población. Si tenemos una población de 15 frascos de un compuesto producido en un día y probamos cada uno para cuantificar su pureza, nuestros datos serán como los de la tabla 4-5. Vimos en la tabla 4-6 cómo usar esos datos para obtener la media (columna 1 dividida entre N = 2.49/15), la desviación de cada valor respecto a la media (columna 3), el cuadrado de la desviación de cada uno respecto a la media (columna 4) y la suma de los cuadrados de las desviaciones. Y con esto podemos calcular la variancia, que es .0034% al cuadrado. (En la Tabla 4-6 también se calcula σ^2 usando la segunda mitad de la ecuación 4-6, $\frac{\Sigma x^2}{N} - \mu^2$. Nótese que conseguimos los mismos resultados pero con un poco menos de trabajo, puesto que no tenemos que calcular las desviaciones respecto a la media.) Al extraer la raíz cuadrada de σ^2 podemos obtener la desviación estándar: .058%.

TABLA 4-5 Resultados de la prueba de pureza hecha en los compuestos

PORCENTAJE OBSERVADO DE IMPUREZA				
.04	.14	.17	.19	.22
.06	.14	.17	.21	.24
.12	.15	.18	.21	.25

Usos de la desviación estándar

Teorema de Chebyshev

La desviación estándar nos permite determinar, con mayor grado de precisión, dónde se sitúan los valores de una distribución de frecuencia en relación con la media. Y esto podemos hacerlo conforme al teorema formulado por el matemático ruso, P. L. Chebyshev (1821-1894). El teorema de Chebyshev establece que, cualquiera que sea la forma de la distribución, por lo menos 75% de los valores caerán dentro de 2 desviaciones estándar positivas y negativas respecto de la media de la distribución, y un mínimo de 89% de los valores se hallará a 3 desviaciones estándar positivas y negativas respecto de la media.

Podemos medir con mayor precisión aún el porcentaje de elementos que caen dentro de intervalos específicos bajo la curva simétrica en forma de campana como la de la figura 4-4. En tales casos podemos afirmar lo siguiente:

1. Cerca de 68% de los valores de la población caerán dentro de 1 desviación estándar más o menos respecto de la media.
2. Cerca de 95% de los valores se encontrarán dentro de 2 desviaciones estándar positivas y negativas respecto de la media.
3. Cerca de 99% de los valores se hallarán en un intervalo que fluctúa entre 3 desviaciones estándar debajo de la media y 3 desviaciones estándar arriba de la media.

Aplicación del teorema de Chebyshev

A la luz del teorema de Chebyshev analizaremos los datos de la tabla 4-6. En ella la impureza media de los 15 frascos del compuesto es .166% y la desviación estándar es .058%. El teorema de Chebyshev señala que por lo menos 75%

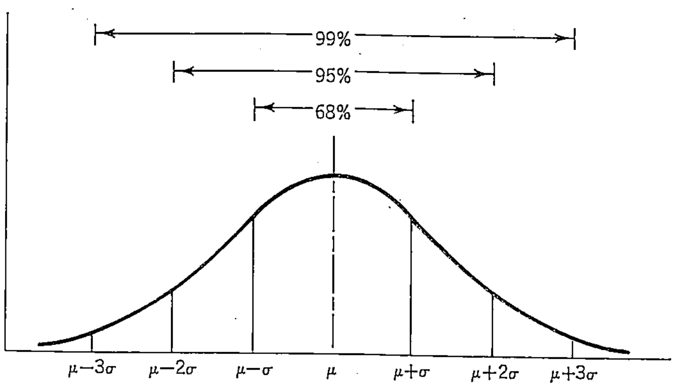


FIGURA 4-4 Localización de las observaciones alrededor de la media en una distribución de frecuencia en forma de campana

de los valores (un mínimo de 11 de nuestros 15 elementos) se encuentran entre .166 − 2(.058) = .050 y .166 + 2(.058) = .282. En efecto, 93% de los valores (14 de los 15 frascos) están realmente en ese intervalo. Nótese que la distribución es bastante simétrica y que 93% está cerca del 95% teórico para un intervalo de dos desviaciones estándar positivas y negativas respecto de la media de una curva en forma de campana.

Concepto de la puntuación estándar

La desviación estándar es también útil para describir cuánto se apartan de la media de la distribución los elementos individuales de esta última. Una medida denominada *puntuación estándar* nos da el número de desviaciones estándar a que determinada observación se encuentra por debajo o encima de la media. Si con x denotamos la observación, la puntuación estándar calculada de los datos de la población será:

$$\text{Puntuación estándar de la población} = \frac{x-\mu}{\sigma} \qquad [4\text{-}8]$$

TABLA 4-6 Determinación de la variancia y la desviación estándar del porcentaje de impureza de los compuestos

OBSERVACION (x) (1)		MEDIA (μ) = 2.49/15 (2)		DESVIACION ($x-\mu$) (3) = (1) − (2)	DESVIACION AL CUADRADO $(x-\mu)^2$ (4) = [(1) − (2)]^2	OBSERVACION AL CUADRADO (x^2) (5) = (1)^2
.04	−	.166	=	−.126	.016	.0016
.06	−	.166	=	−.106	.011	.0036
.12	−	.166	=	−.046	.002	.0144
.14	−	.166	=	−.026	.001	.0196
.14	−	.166	=	−.026	.001	.0196
.15	−	.166	=	−.016	.000	.0225
.17	−	.166	=	.004	.000	.0289
.17	−	.166	=	.004	.000	.0289
.18	−	.166	=	.014	.000	.0324
.19	−	.166	=	.024	.001	.0361
.21	−	.166	=	.044	.002	.0441
.21	−	.166	=	.044	.002	.0441
.22	−	.166	=	.054	.003	.0484
.24	−	.166	=	.074	.005	.0576
.25	−	.166	=	.084	.007	.0625
2.49 ← Σx					.051 ← $\Sigma(x-\mu)^2$	.4643 ← Σx^2

$$\sigma^2 = \frac{\Sigma(x-\mu)^2}{N} \qquad [4\text{-}6]$$
$$= \frac{.051}{15}$$
$$= .0034 \text{ porcentaje al cuadrado}$$

← O BIEN →

$$\sigma^2 = \frac{\Sigma x^2}{N} - \mu^2 \qquad [4\text{-}6]$$
$$= \frac{.4643}{15} - (.166)^2$$
$$= .0034 \text{ porcentaje al cuadrado}$$

$$\sigma = \sqrt{\sigma^2} \qquad [4\text{-}7]$$
$$= \sqrt{.0034}$$
$$= .058\ \%$$

donde

- x = la observación de la población
- μ = la media de la población
- σ = la desviación estándar de la población

Supóngase que observamos un frasco de un compuesto que tiene una impureza de .108%. Dado que nuestra población tiene una media de .166 y una desviación estándar de .058, una observación de .108 tendrá una puntuación estándar de −1:

Cálculo de la puntuación estándar

$$\text{Puntuación estándar} = \frac{x-\mu}{\sigma} \quad [4\text{-}8]$$

$$= \frac{.108 - .166}{.058}$$

$$= -\frac{.058}{.058}$$

$$= -1$$

La impureza observada de .282% tendrá una puntuación estándar de +2:

$$\text{Puntuación estándar} = \frac{x-\mu}{\sigma} \quad [4\text{-}8]$$

$$= \frac{.282 - .166}{.058}$$

$$= \frac{.116}{.058}$$

$$= 2$$

Interpretación de la puntuación estándar

La puntuación estándar indica que una impureza de .282% se desvía de la media en 2(.058) = .116 unidades, lo cual es igual a +2 en términos de las unidades de desviación estándar respecto a la media.

Cálculo de la variancia y de la desviación estándar usando datos agrupados

Cálculo de la variancia y de la desviación estándar en datos agrupados

En el ejemplo con que empieza el capítulo, los datos sobre las ventas en 100 restaurantes de comida de preparación rápida ya estaban agrupados en una distribución de frecuencia. A tales datos podemos aplicar las siguientes fórmulas para calcular la variancia y la desviación estándar:

$$\sigma^2 = \frac{\Sigma f(x-\mu)^2}{N} = \frac{\Sigma f x^2}{N} - \mu^2 \quad [4\text{-}9]$$

y

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\Sigma f(x-\mu)^2}{N}} = \sqrt{\frac{\Sigma f x^2}{N} - \mu^2} \quad [4\text{-}10]$$

donde:

- σ^2 = variancia de la población
- σ = desviación estándar de la población
- f = frecuencia de cada una de las clases
- x = marca de clase de las clases
- μ = media de la población
- N = tamaño de la población

La tabla 4-7 muestra cómo aplicar estas ecuaciones para obtener la variancia y la desviación estándar de las ventas en 100 restaurantes de comida de preparación rápida.

Podemos dejar como ejercicio para el lector interesado en los cálculos verificar que la segunda mitad de la ecuación 4-9, $\frac{\Sigma f x^2}{N} - \mu^2$, produzca el mismo valor de σ^2.

Transición a la variancia y a la desviación estándar de la muestra

Ahora ya podemos calcular los estadísticos muestrales que son análogos a la variancia de la población σ^2 y la desviación estándar de la población σ. Son la variancia muestral s^2 y la desviación estándar de la muestra s. En la siguiente sección, el lector advertirá que estamos haciendo la transición de las letras griegas (que denotan los parámetros de la población) a las letras latinas de los estadísticos muestrales.

Desviación estándar de la muestra

Cálculo de la desviación estándar de la muestra

Para obtener la variancia y la desviación estándar de la muestra, aplicamos las mismas fórmulas que las ecuaciones 4-6 y 4-7, sustituyendo μ por $\bar{x}$ y N por $n-1$. He aquí las fórmulas:

$$s^2 = \frac{\Sigma(x-\bar{x})^2}{n-1} = \frac{\Sigma x^2}{n-1} - \frac{n\bar{x}^2}{n-1} \quad [4\text{-}11]$$

y

$$s = \sqrt{s^2} = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}} = \sqrt{\frac{\Sigma x^2}{n-1} - \frac{n\bar{x}^2}{n-1}} \quad [4\text{-}12]$$

donde:

- s^2 = variancia de la muestra
- s = desviación estándar de la muestra
- x = valor de cada una de n observaciones
- $\bar{x}$ = media de la muestra
- $n-1$ = número de observaciones de la muestra −1

TABLA 4-7 Determinación de la variancia y la desviación estándar de las ventas en 100 restaurantes de alimentos de preparación rápida en el distrito oriental (en miles)

CLASE	MARCA DE CLASE (x) (1)	FRECUENCIA (f) (2)	$f \times x$ (3) = (2) × (1)	MEDIA (μ) (4)	$x - \mu$ (1) − (4)	$(x-\mu)^2$ [(1) − (4)]²	$f(x-\mu)^2$ (2) × [(1) − (4)]²
700– 799	750	4	3,000	1,250	−500	250,000	1,000,000
800– 899	850	7	5,950	1,250	−400	160,000	1,120,000
900– 999	950	8	7,600	1,250	−300	90,000	720,000
1,000–1,099	1,050	10	10,500	1,250	−200	40,000	400,000
1,100–1,199	1,150	12	13,800	1,250	−100	10,000	120,000
1,200–1,299	1,250	17	21,250	1,250	0	0	0
1,300–1,399	1,350	13	17,550	1,250	100	10,000	130,000
1,400–1,499	1,450	10	14,500	1,250	200	40,000	400,000
1,500–1,599	1,550	9	13,950	1,250	300	90,000	810,000
1,600–1,699	1,650	7	11,550	1,250	400	160,000	1,120,000
1,700–1,799	1,750	2	3,500	1,250	500	250,000	500,000
1,800–1,899	1,850	1	1,850	1,250	600	360,000	360,000
		100	125,000				6,680,000

$$\bar{x} = \frac{\Sigma(f \times x)}{n} \qquad [3\text{-}3]$$

$$= \frac{125,000}{100}$$

$$= 1,250 \text{ dólares} \leftarrow \text{media}$$

$$\sigma^2 = \frac{\Sigma f(x-\mu)^2}{N} \qquad [4\text{-}9]$$

$$= \frac{6,680,000}{100}$$

$$= 66,800 \text{ (o 66,800 dólares al cuadrado)} \leftarrow \text{variancia}$$

$$\sigma = \sqrt{\sigma^2} \qquad [4\text{-}10]$$

$$= \sqrt{66,800}$$

$$= 258.5 \leftarrow \text{desviación estándar} = \$258,500$$

Uso de $n - 1$ como denominador

¿Por qué usamos como denominador $n - 1$ y no n? Los estadísticos pueden probar que, si extraemos muchas muestras de determinada población, si obtenemos la variancia de la muestra (s^2) para cada muestra y si promediamos todo esto, el promedio tiende a no ser igual a la variancia de la población, σ^2, a menos que utilicemos como denominador $n - 1$. En el capítulo 8, aprenderemos la explicación estadística de por qué esto es así.

Cálculo de la variancia y la desviación estándar de la muestra para los datos del Cumberland Hospital

Las ecuaciones 4-11 y 4-12 nos permiten calcular la variancia y la desviación estándar de la muestra de los pagos hechos por Blue Cross-Blue Shield al hospital Cumberland, mencionados en la tabla 4-4 de la página 129. Esto lo hacemos en la tabla 4-8, puntualizando que ambas mitades de la ecuación 4-11 producen el mismo resultado.

TABLA 4-8 Determinación de la variancia y la desviación estándar de los pagos anuales de Blue Cross-Blue Shield al Hospital Cumberland (en miles)

OBSERVACION (x) (1)	MEDIA ($\bar{x}$) (2)	$x - \bar{x}$ (1) − (2)	$(x-\bar{x})^2$ [(1) − (2)]²	x^2 (1)²
863	1,351	−488	238,144	744,769
903	1,351	−448	200,704	815,409
957	1,351	−394	155,236	915,849
1,041	1,351	−310	96,100	1,083,681
1,138	1,351	−213	45,369	1,295,044
1,204	1,351	−147	21,609	1,449,616
1,354	1,351	3	9	1,833,316
1,624	1,351	273	74,529	2,637,376
1,698	1,351	347	120,409	2,883,204
1,745	1,351	394	155,236	3,045,025
1,802	1,351	451	203,401	3,247,204
1,883	1,351	532	283,024	3,545,689
			$\Sigma(x-\bar{x})^2 \rightarrow$ 1,593,770	23,496,182 $\leftarrow \Sigma x^2$

$$s^2 = \frac{\Sigma(x-\bar{x})^2}{n-1} \qquad [4\text{-}11]$$

$$= \frac{1,593,770}{11}$$

$$= 144,888 \text{ (o \$144,888 millones al cuadrado)} \leftarrow \text{variancia muestral}$$

$$s = \sqrt{s^2} \qquad [4\text{-}12]$$

$$= \sqrt{144,888}$$

$$= 380.64 \text{ (es decir, \$380,640)} \leftarrow \text{desviación estándar de la muestra}$$

O

$$s^2 = \frac{\Sigma x^2}{n-1} - \frac{n\bar{x}^2}{n-1} \qquad [4\text{-}11]$$

$$= \frac{23,496,182}{11} - \frac{12(1,351)^2}{11}$$

$$= \frac{1,593,770}{11}$$

$$= 144,888$$

Cálculo de la muestra y las puntuaciones estándar

Del mismo modo que hemos utilizado la desviación estándar de la población para derivar las puntuaciones estándar de la población, también podemos servirnos de la desviación de la muestra para obtener las puntuaciones estándar de la muestra. Dichas puntuaciones nos indican cuántas desviaciones estándar una observación particular de la muestra se encuentra por encima o debajo de la media muestral. La fórmula apropiada es:

$$\text{Puntuación estándar de la muestra} = \frac{x - \bar{x}}{s} \qquad [4\text{-}13]$$

donde:

- x = la observación de la muestra
- $\bar{x}$ = la media muestral
- s = la desviación estándar de la muestra

En el ejemplo que acabamos de presentar, vemos que la observación 863 corresponde a una puntuación estándar de -1.28:

$$\text{Puntuación estándar de la muestra} = \frac{x-\bar{x}}{s} \quad [4\text{-}13]$$

$$= \frac{863-1{,}351}{380.64}$$

$$= \frac{-488}{380.64}$$

$$= -1.28$$

Características de la desviación estándar

En esta sección hemos demostrado por qué la desviación estándar es la medida de dispersión de mayor uso. Con ella podemos comparar las distribuciones y obtener las puntuaciones estándar, aspecto muy importante de la inferencia estadística que veremos después. A semejanza de la desviación absoluta promedio, incluye todas las observaciones del conjunto de datos. Pero también tiene sus desventajas. No es tan fácil calcular el intervalo, y tampoco puede obtenerse en las distribuciones abiertas. Además, los valores extremos presentes en los datos distorsionan el valor de la desviación estándar, aunque en grado menor que el intervalo.

EJERCICIOS

4-17 Una empresa hollywoodense de repartos, está escogiendo un grupo de extras para una película. Las edades de los primeros 20 varones entrevistados son:

50	56	55	49	52	57	56	57	56	59
54	55	61	60	51	59	62	52	54	49

El director de la película quiere hombres cuya edad se agrupe estrechamente alrededor de los 55 años. Como es un entusiasta aficionado de la estadística, sugiere que una desviación estándar de 3 años sea aceptable. ¿Cumple con el requisito este grupo de extras?

4-18 A continuación se transcriben los datos de una muestra de la producción diaria de embarcaciones de fibra de vidrio en Hydrosport, Ltd., un armador de Miami:

16	20	17	26	18	22	21	23	19	24

El gerente de producción de la compañía piensa que cualquier desviación absoluta promedio de más de 3 embarcaciones por día indica una variación inaceptable en la tasa de producción. ¿Deben inquietarle las tasas de producción de la planta?

4-19 En un conjunto de 60 observaciones con una media de 66.8, una variancia de 12.6 y una forma desconocida de la distribución:

a) ¿Entre qué valores deberá caer por lo menos 75% de las observaciones según el teorema de Chebyshev?
b) Si la distribución es simétrica y tiene forma de campana, ¿aproximadamente cuántas observaciones se encontrarán necesariamente en el intervalo 59.7 a 73.9?
c) Encuentre las puntuaciones estándar de las siguientes observaciones en la distribución: 61.45, 75.37, 84.65 y 51.50.

4-20 El número de cheques cobrados diariamente en 5 sucursales de un banco durante el mes anterior tuvo la siguiente distribución de frecuencia:

Clase	Frecuencia
0–199	10
200–399	13
400–599	17
600–799	42
800–999	18

El director de operaciones del banco, sabe que una desviación estándar en el cobro de los cheques de más de 200 cheques diarios crea problemas de organización y dotación de personal en las sucursales, debido a una carga de trabajo no uniforme. ¿Debe preocuparse en este momento?

4-21 El Federal Reserve Board ha concedido permiso a todos los bancos miembros para elevar las tasas de interés en ½% a todos los depositantes. Las tasas anteriores que se pagaban a los ahorros eran de 5¼%; para los certificados de depósito: a un año, 7½%; a 18 meses, 8¾%; a 2 años, 9%; a 3 años, 10½%; a 5 años, 11%. El presidente del First State Bank quiere saber cuáles serán las características de la nueva distribución de las tasas, si se agrega ½% a todas ellas. ¿Qué relación guardan las nuevas características con las anteriores?

4-22 El administrador de un hospital de Georgia realizó una encuesta sobre el número de días que 200 pacientes escogidos al azar permanecen en él después de ser sometidos a una operación. He aquí los datos:

Estancia en el hospital (en días)	1-3	4-6	7-9	10-12	13-15	16-18	19-21	22-24
Frecuencia	24	83	52	22	11	5	2	1

a) Calcule la media y la desviación estándar.
b) Según el teorema de Chebyshev, ¿cuántas estancias en el hospital fluctuarán entre 0 y 15 días? ¿Cuántas se hallan realmente en ese intervalo?
c) Puesto que esta distribución tiene aproximadamente la forma de campana, ¿cuántas estancias cabe esperar que oscilen entre 0 y 15 días?

4-23 Con el propósito de estimar la demanda futura, la National Motor Company efectuó un estudio en que preguntó a un grupo de matrimonios cuántos automóviles deberán poseer en 1990 los que deseen economizar combustible. En cada matrimonio, la compañía sacó el promedio de las respuestas de ambos cónyuges para obtener la respuesta global del matrimonio. Las respuestas fueron después tabuladas en una distribución de frecuencia.

Número de autos	0-.49	.50-.99	1.00-1.49	1.50-1.99	2.00-2.49	2.50-2.99
Frecuencia	2	14	23	7	4	2

a) Calcule la variancia y la desviación estándar.
b) Dado que la distribución tiene forma de campana, ¿cuántas observaciones deben caer, en teoría, entre .7 y 1.8? ¿Entre .2 y .24? ¿Cuántas caen realmente en esos intervalos?

4-24 El propietario de una panadería dijo que la producción semanal promedio de su compañía fue de 11,398 hogazas, con una variancia de 49,729. Si los datos usados para calcular los resultados fueron recogidos durante 32 semanas, ¿en cuántas semanas estuvo el nivel de producción por debajo de 11,175? ¿Y por arriba de 11,844?

4-25 Una compañía publicitaria norteamericana cuenta con 3 oficinas en 3 ciudades diferentes. Las tarifas de los sueldos difieren de un estado a otro. En la oficina de Washington, D.C., el incremento promedio de sueldos fue de $1,500 el año pasado, con una desviación estándar de $400. En la oficina de Nueva York, el aumento promedio fue de $3,760, con una desviación estándar de $622. En Durham, N.C., el incremento promedio fue de $850, con una desviación estándar de $95. Se entrevistó a tres empleados. El de Washington recibió un aumento de $1,100; el de Nueva York un incremento de $3,200; y el de Durham, un incremento de $500. ¿Cuál de los 3 obtuvo el menor aumento en relación con la media y la desviación estándar de esta oficina?

4-26 Una fábrica de alimentos comercializa ampliamente 3 diferentes productos a nivel nacional. Uno de los objetivos de los anuncios de los productos consiste en lograr que el consumidor sepa que esta empresa es la que los fabrica. Para medir la eficacia con que el anuncio crea el reconocimiento, se pidió a un grupo de consumidores identificar lo más rápidamente posible la compañía que elabora cada uno de una larga lista de productos. El primer producto de la empresa tuvo una latencia promedio de respuesta de 2 segundos, con una desviación estándar de .005 segundos. El segundo tuvo una latencia promedio de 3 segundos, con una desviación estándar de .007. El tercero tuvo un periodo de latencia de 4 segundos, con una desviación estándar de .10 segundos. Un sujeto particular mostró las siguientes latencias: 1.994 en el primer producto, 3.007 en el segundo y 3.990 en el tercero. ¿En qué producto este sujeto estuvo más lejos que los demás del desempeño promedio, en unidades de desviación estándar?

4-27 Antonio Balderrama es un médico que se especializa en el conocimiento y en el empleo eficaz de analgésicos destinados a enfermos graves. A fin de conocer aproximadamente cuántas enfermeras y personal debe emplear en su consultorio, ha empezado a llevar un control del número de pacientes a quienes atiende a la semana. Cada semana anota el número de enfermos graves y el de pacientes normales. El señor Balderrama tiene motivos para pensar que el número de pacientes normales por semana presentaría el aspecto de una curva en forma de campana si dispusiera de suficientes datos. (Esto no sucede en el caso de los pacientes graves.) Sin embargo, cuenta con datos sólo para las últimas 5 semanas.

Pacientes graves	33	50	22	27	48
Pacientes ordinarios	34	31	37	36	27

a) Calcule la media y la variancia para el número de enfermos graves por semana. Use el teorema de Chebyshev para encontrar los límites dentro de los cuales caerá el "75% intermedio" de los números de enfermos graves a quienes el médico atienda a la semana.
b) Calcule la media, la variancia y la desviación estándar del número de pacientes normales por semana. ¿Dentro de qué límites deberá caer el "68% intermedio" de esas cifras semanales?

4-28 El superintendente de cualquier distrito escolar de la localidad afronta siempre dos problemas muy serios. El primero de ellos es tratar con el consejo escolar recién elegido y el segundo es la necesidad de estar preparado para buscar un nuevo trabajo, debido al primer problema. Tomás Lemus, superintendente del distrito escolar 18, no es la excepción a esta regla. Ya aprendió el valor de conocer todas las cifras de un presupuesto y sabe aprovecharlas para beneficio personal. Este año, el consejo escolar recomendó un presupuesto de $350,000 para la investigación de los medios de comunicación. Por experiencia, Tomás sabe que los gastos siempre rebasan

el presupuesto, y la cantidad en que lo superan tiene una media de $40,000 y una variancia de $100,000,000 de dólares al cuadrado. Tomás se enteró de la existencia del teorema de Chebyshev en la universidad, y a su juicio éste podría ayudarle a obtener un intervalo de valores donde caerían los gastos reales el 75% de las veces en los años en que el presupuesto sea igual al de este año. Hágale un favor a Tomás y calcule ese intervalo.

4-29 Una famosa psicóloga clínica, tiene expedientes muy completos sobre todos sus pacientes. Con esos datos ha desarrollado 4 categorías donde los coloca: niños, adultos jóvenes, adultos y ancianos. En cada categoría ha calculado el coeficiente intelectual (CI) promedio y su variancia dentro de esa categoría. Estas cifras se dan en la tabla anexa. Si en cierto día la psicóloga atendió a 4 pacientes (uno de cada categoría) cuyos cocientes intelectuales fueron los siguientes: niño, 90; adulto joven, 92; adulto, 100; anciano, 98; ¿cuál de ellos tuvo el coeficiente intelectual más lejano de la media, en unidades de desviaciones estándar, para esa categoría particular?

CATEGORIA	CI MEDIO	VARIANCIA DEL CI
Niño	110	81
Adulto joven	90	64
Adulto	95	49
Anciano	90	121

4-4 DISPERSION RELATIVA: el coeficiente de variación

La desviación estándar es una medida *absoluta* de dispersión y expresa la variación en las mismas unidades que los datos originales. Los pagos anuales de Blue Cross-Blue Shield al hospital Cumberland (Tabla 4-8) tienen una desviación estándar de $380,640. Y los reembolsos que hacen al Valley Falls Hospital (Tabla 4-1) muestran una desviación estándar (que el lector calcula) de $57,390. ¿Podemos comparar los valores de esas dos desviaciones estándar? Por desgracia, la respuesta es negativa.

Desventajas de la desviación estándar

La desviación estándar no puede ser la única base para comparar dos distribuciones. Si tenemos una desviación estándar de 10 y una media de 5, los valores varían en una cantidad que es el doble de la media. En cambio, si tenemos una desviación estándar de 10 y una media de 5,000, la variación relativa a la media es insignificante. Por tanto, no podemos conocer la dispersión de un conjunto de datos sin conocer antes la desviación estándar, la media y la relación de tamaño existente entre ésta y la desviación estándar.

El coeficiente de variación, una medida relativa

Lo que necesitamos es una medida *relativa* que nos dé una idea general de la magnitud de la desviación estándar en relación con la magnitud de la media. El *coeficiente de variación* (cv) es precisamente esa medida de dispersión. Relaciona una y otra expresando la desviación estándar como porcentaje de la media. De ahí que la unidad de medida se llame "por ciento" en vez de las mismas unidades que los datos originales.

En una población, la fórmula del coeficiente de variación es:

$$\text{Coeficiente de variación de la población} = \frac{\overbrace{\sigma}^{\text{Desviación estándar de la población}}}{\underbrace{\mu}_{\text{Media de la población}}}(100) \qquad [4\text{-}14]$$

SAS
CALIFICACIONES DE 1985 EN ESTADISTICA PARA LOS
ESTADISTICOS RESUMIDOS DEL CURSO COMPLETO POR SECCION

VARIABLE	N	MEDIA	DESVIACION ESTANDAR	VARIANCIA	INTERVALO	C.V.
EXAM1	199	50.221106	9.4885022	90.031674	52.000000	18.893
EXAM2	199	56.894472	10.7126810	114.761535	57.000000	18.829
TAREA	199	108.597990	19.0127335	361.484036	122.000000	17.507
FINAL	199	45.281407	10.0141923	100.284046	61.000000	22.115
TOTAL	199	68.567959	11.2397325	126.331586	76.103163	16.392
SECCION = 1						
EXAM1	27	47.148148	10.8617314	117.977208	48.000000	23.037
EXAM2	27	53.296296	13.5868002	184.601140	52.000000	25.493
TAREA	27	109.074074	20.5106463	420.686610	102.000000	18.804
FINAL	27	45.740741	10.6792127	114.045584	50.000000	23.347
TOTAL	27	67.103469	13.6211984	185.537045	65.043309	20.299
SECCION = 2						
EXAM1	46	50.826087	10.6129990	112.635749	43.0000000	20.881
EXAM2	46	58.260870	10.8370451	117.441546	49.0000000	18.601
TAREA	46	112.521739	17.6417927	311.232850	79.0000000	15.679
FINAL	46	44.760870	11.8999249	141.608213	61.0000000	26.586
TOTAL	46	69.388850	12.4966506	156.166277	60.3211679	18.010
SECCION = 3						
EXAM1	37	53.189189	8.9809608	80.657658	33.0000000	16.885
EXAM2	37	60.513514	7.6034408	57.812312	28.0000000	12.565
TAREA	37	111.783784	16.8030009	282.340841	96.0000000	15.032
FINAL	37	49.081081	7.3650012	54.243243	29.0000000	15.006
TOTAL	37	72.816098	8.8560862	78.430262	34.8340633	12.162
SECCION = 4						
EXAM1	26	50.769231	8.7466917	76.504615	37.0000000	17.228
EXAM2	26	59.384615	6.4440790	41.526154	27.0000000	10.851
TAREA	26	104.576923	15.0417368	226.253846	65.0000000	14.383
FINAL	26	44.923077	8.0643565	65.033846	31.0000000	17.951
TOTAL	26	68.600037	8.0832528	65.338975	32.0038929	11.783
SECCION = 5						
EXAM1	36	49.472222	8.1643340	66.656349	37.000000	16.503
EXAM2	36	55.944444	11.4391419	130.853968	47.000000	20.447
TAREA	36	107.361111	24.3359262	592.237302	120.000000	22.667
FINAL	36	44.333333	10.3730420	107.600000	40.000000	23.398
TOTAL	36	67.428710	11.8180624	139.666600	51.431630	17.527
SECCION = 6						
EXAM1	27	48.666667	8.4352739	71.153846	32.0000000	17.333
EXAM2	27	52.074074	11.0902796	122.994302	35.0000000	21.297
TAREA	27	102.592593	17.0322302	290.096866	53.0000000	16.602
FINAL	27	42.111111	9.4353400	89.025641	38.0000000	22.406
TOTAL	27	64.300478	9.8485111	96.993170	35.9532847	15.316

FIGURA 4-5 Salida de un programa SAS, que muestra los estadísticos resumidos de los datos referentes a las calificaciones obtenidas en los cursos

Aplicando esta fórmula en un ejemplo, supondremos que cada día el técnico de laboratorio A realiza 40 análisis, con una desviación estándar de 5. El técnico B termina 160 análisis diarios, con una desviación estándar de 15. ¿Cuál de los dos muestra menor variabilidad?

A primera vista parecería que el técnico B tiene el triple de variación en la tasa de salida que el técnico A. Pero el primero realiza los análisis a una velocidad cuatro veces mayor. Para tener en cuenta toda esta información, calcule el coeficiente de variación de ambos técnicos:

Cálculo del coeficiente de variación

$$\text{Coeficiente de variación} = \frac{\sigma}{\mu}(100) \qquad [4\text{-}14]$$

$$= \frac{5}{40}(100)$$

$$= 12.5\% \leftarrow \text{para el técnico A}$$

y:

$$\text{Coeficiente de variación} = \frac{15}{160}(100)$$

$$= 9.4\% \leftarrow \text{para el técnico B}$$

Y así descubrimos que el técnico B, quien tiene una mayor variación *absoluta* en la salida que el técnico A, muestra una menor variación *relativa*, ya que la salida media de B es mucho mayor que la suya.

Uso de la computadora para obtener las medidas de la tendencia central y de variabilidad

En el caso de extensos conjuntos de datos, utilizamos la computadora para calcular nuestra medida de la tendencia central y de la variabilidad. En la figura 4-5, hemos aplicado el sistema SAS para obtener algunos de esos estadísticos resumidos relativos a los datos de calificaciones en el apéndice 9. Se muestran los estadísticos de cada sección, lo mismo que los del curso en su totalidad.

EJERCICIOS

4-30 Los pesos del equipo profesional de fútbol americano Baltimore Bullets tienen una media de 224 libras, con una desviación estándar de 18 libras, mientras que el peso medio y la desviación estándar de su oponente, los Chicago Trailblazers, son 195 y 12, respectivamente. ¿Cuál de los dos equipos muestra la mayor dispersión relativa en el peso de los miembros del equipo?

4-31 Electrónica Moderna es una compañía que está considerando la conveniencia de implantar 2 programas de capacitación. A dos grupos se les impartió capacitación para realizar la misma tarea. El grupo 1 fue adiestrado con el programa A; el grupo 2, con el programa B. En el primer grupo, se requirieron en promedio 32.11 horas para capacitar a cada empleado, con una variancia de 68.09. En el segundo grupo, se necesitó un promedio de 19.75 horas para capacitar a cada empleado, con una variancia de 71.14. ¿Cuál programa mostró la menor variabilidad relativa en sus resultados?

4-32 Con las dos siguientes muestras se describen las muestras de las edades de los estudiantes que asisten al programa diurno y al programa nocturno de posgrado en administración en la Universidad Central:

Programa diurno de administración	24	30	28	23	25	22	26	27	28	25
Programa nocturno de administración	26	33	29	28	27	29	33	34	27	28

Si la homogeneidad del grupo escolar es un factor positivo del aprendizaje, aplique una medida de variabilidad relativa que indique a cuál de los dos grupos será más fácil enseñarle.

4-33 Hay varias medidas posibles del desempeño de ventas, entre ellas la constancia con que un vendedor cumple con las metas establecidas. Los datos siguientes representan el porcentaje de la meta lograda por 3 vendedores en los últimos 5 años.

Patricia	88	68	89	92	103
Juan	76	88	90	86	79
Francisco	104	88	118	88	123

a) ¿Cuál vendedor es el más constante?
b) Comente la conveniencia de aplicar una medida de constancia junto con el porcentaje de la obtención de metas en ventas para evaluar el desempeño de un vendedor.
c) ¿Puede recomendar otra medida de constancia, en el rendimiento, que sea más adecuada para este caso?

4-34 El consejo de administración de una corporación está estudiando la posibilidad de adquirir una de dos compañías y con mucho detenimiento analiza la administración de cada una en relación con su inclinación a correr riesgos. En los últimos 5 años, la primera compañía alcanzó un promedio de rendimiento sobre la inversión de 28%, con una desviación estándar de 5.3%. La segunda tuvo un rendimiento promedio de 37.8%, con una desviación estándar de 4.8%. Si suponemos que el riesgo se acompaña de una mayor dispersión relativa, ¿cuál de estas dos empresas ha aplicado una estrategia más riesgosa?

4-35 Una compañía farmacéutica que suministra a los hospitales ciertos medicamentos ya dosificados usa varias máquinas para producir fármacos que requieren diferentes dosis. Una máquina, diseñada para producir dosis de 100 cc, tiene como dosis media 100 cc, con una desviación estándar de 5.2 cc. Otra produce cantidades de 180 cc de medicamento ya dosificado y muestra una desviación estándar de 8.6 cc. ¿Cuál de las máquinas tiene la menor exactitud desde el punto de vista de la dispersión relativa?

4-36 Una compañía mayorista estaba estudiando la posibilidad de convertirse en proveedor de 3 minoristas, pero la escasez de inventario la obligó a seleccionar un solo minorista. El gerente de crédito de la compañía está evaluando los créditos de los tres. En los últimos 5 años, sus cuentas por cobrar se han atrasado el siguiente número promedio de días. El gerente de crédito considera que la consistencia, además de un promedio mínimo, es de suma importancia. Basándose en la dispersión relativa, ¿cuál minorista será el mejor cliente?

Hernández	62.2	61.8	63.4	63.0	61.7
Luján	62.5	61.9	62.8	63.0	60.7
Díaz	62.0	61.9	63.0	63.9	61.5

4-37 Una compañía vende 3 grados de semilla que se distingue por la consistencia de la germinación de las semillas. El laboratorio estatal de pruebas de semillas tiene una muestra de cada grado; a continuación se incluyen los resultados de las pruebas practicadas a varias semillas que germinaron y que provenían de paquetes de 100:

Grado I (regular)	88	91	92	89	79
Grado II (extra)	87	92	88	90	92
Grado III (super)	90	89	79	93	88

¿Es lógica la graduación de las semillas de la empresa?

4-38 Una fábrica de aparatos de uso doméstico acaba de terminar un estudio sobre tres posibles configuraciones de líneas de montaje para producir las tostadoras de dos rebanadas que más vende. La configuración 1 ha aportado un tiempo medio de fabricación de tostadoras de 36.7 minutos, con una desviación estándar de 5 minutos. La configuración II ha arrojado una media de 24.2 minutos, con una desviación estándar de 8 minutos. La configuración III ha producido una media de 38.3 minutos, con una desviación estándar de 4 minutos. ¿Cuál configuración de la línea de montaje tiene la menor variación relativa en el tiempo que se necesita para fabricar una tostadora?

4-5 ANALISIS EXPLORATORIO DE DATOS

Suposiciones necesarias en el análisis clásico

En los capítulos 2, 3 y 4 nos hemos ocupado de la *presentación* de los datos: ¿cómo se han de organizar y sintetizar los datos brutos de manera que reconozcamos las características más importantes de ellos. El resto del libro está dedicado casi enteramente a los métodos clásicos del análisis estadístico de los datos que se emplean una vez reunidos y organizados los datos. Como veremos al hablar de dichos métodos, muchos de los análisis clásicos se fundan en las suposiciones que es preciso hacer sobre los datos que van a ser analizados.

Métodos robustos de análisis

En los trabajos recientes, presididos principalmente por el profesor John W. Tukey de Princeton University y de Bell Telephone Laboratories, se ha intentado idear métodos para analizar los datos que exigen poquísimas suposiciones previas. A esos métodos los estadísticos los llaman *robustos*. Tales técnicas de *análisis exploratorio* de datos (EDA) permiten al estadístico examinar los datos y determinar qué análisis ulterior puede resultar idóneo.

Alternativas para realizar el análisis exploratorio

Los paquetes de software de mayor uso en la realización del análisis estadístico tienen la capacidad de ejecutar el análisis exploratorio de datos. La figura 4-6 contiene la salida cuando se usa el paquete SAS para efectuar un análisis elemental sobre los datos referentes a la concentración de cloro que se incluyeron en el capítulo 2. Examinaremos brevemente esa salida; si el lector desea información más completa sobre el análisis exploratorio de datos, le recordamos que la bibliografía que viene al final del libro ofrece varias obras y artículos sobre el tema.

Cuartiles, intervalos y percentiles

La primera sección de la salida (titulada "momentos") proporciona la media, la desviación estándar y las medidas numéricas de sesgo y curtosis de los datos. Como ya vimos en los capítulos 2-4, tales cantidades nos indican la *forma* de los datos.

Representaciones gráficas de los datos

La siguiente sección de la salida (titulada "cuantiles") contiene los cuartiles y varios intervalos, así como algunos percentiles que delinean los extremos superior (99%, 95%, 90%) e inferior (10%, 5%, 1%) de los datos. Así pues, el análisis exploratorio de datos (EDA) no sólo identifica el centro de los datos; también centra la atención en los valores no centrales y atípicos. A menudo, con

FIGURA 4-6 Un análisis exploratorio de los datos referentes a la concentración de cloro en el capítulo 2, utilizando el paquete de computación SAS

```
      EJEMPLO DEL EMPLEO DE SAS EN EL ANALISIS EXPLORATORIO DE DATOS

                               UNIVARIADO

VARIABLE=PPM            CONCENTRACIONES DE CLORO EN PPM

                               MOMENTOS

          N                    30      SUMA DE PESOS          30
          MEDIA           16.0367      SUMA                481.1
          DESVIACION EST. .411459      VARIANCIA        0.169299
          SESGO          0.345475      CURTOSIS         -0.10233
          USS             7720.15      CSS               4.90967
          CV              2.56574      MEDIA ESTAN     0.0751219
          T:MEDIA=0       213.475      PROB > |T|         0.0001
          INTERVALO SGN     232.5      PROB > |S|         0.0001
          NUM ~= 0             30
          W:NORMAL       0.969853      PROB > W            0.571

                            CUANTILES(DEF = 4)

          100% MAX           16.9          99%            16.9
           75% Q3            16.3          95%          16.845
           50% MIN             16          90%           16.78
           25% Q1          15.775          10%            15.6
            0% MIN           15.2           5%           15.31
                                            1%            15.2
          INTERVALO           1.7
          Q3-Q1          0.524988
          MODA               15.9

                                EXTREMOS

                         MINIMO             MAXIMO
                          15.2               16.4
                          15.4               16.6
                          15.6               16.8
                          15.6               16.8
                          15.6               16.9

        TALLO  HOJA                       #     DIAGRAMA RECTANGULAR

         168 000                          3             |
         166 0                            1             |
         164 00                           2             |
         162 00000                        5         +-------+
         160 00000                        5         *---+---*
         158 0000000                      7         |       |
         156 00000                        5         +-------+
         154 0                            1             |
         152 0                            1             |
         150
             ----+----+----+----+
         MULTIPLICAR TALLO, HOJA POR 10**-01
```

un examen más detenido de estos "últimos" descubriremos que realmente no pertenecen al conjunto de datos. (Quizá se registraron incorrectamente.) Ya hemos visto como esos elementos distorsionan las medias muestrales.

El paquete SAS ofrece varios diagramas de los datos. Los "diagramas de tallos y hojas" se parecen a los histogramas; sólo que muestran simultáneamente todos los valores de datos mientras los agrupan. Por tanto, tienen la ventaja del histograma consistente en resumir los datos, pero sin la desventaja de dejar afuera algunos detalles. Los "diagramas rectangulares" ofrecen una representación gráfica de la mediana (la línea horizontal media en la Tabla 4-9), de los cuartiles (las líneas de la parte superior e inferior de la casilla de la Tabla 4-9) y de los extremos (las líneas punteadas que salen de la casilla). Quizá el lector considere el diagrama rectangular como una distribución esquelética de frecuencia.

4-6 GLOSARIO DEL CAPITULO

ANALISIS EXPLORATORIO DE DATOS (EDA) Métodos con que analizan datos que requieren poquísimas suposiciones previas.

COEFICIENTE DE VARIACION Medida relativa de dispersión, semejante entre las distribuciones, que expresa la desviación estándar en un porcentaje de la media.

CUARTILES Fractiles que dividen los datos en 4 partes iguales.

DECILES Fractiles que dividen los datos en 10 partes iguales.

DESVIACION ABSOLUTA PROMEDIO En un conjunto de datos, la distancia promedio entre las observaciones y la media.

DESVIACION DE CUARTIL Mitad del intervalo del intercuartil; una medida del intervalo promedio de una cuarta parte de los datos.

DESVIACION ESTANDAR Raíz cuadrada positiva de la variancia; una medida de la dispersión, expresada en las mismas unidades que los datos originales y no en las unidades cuadradas de la variancia.

DISPERSION Esparcimiento o variabilidad en un conjunto de datos.

FRACTIL En una distribución de frecuencia, la localización de un valor en determinada fracción de los datos o arriba de ellos.

INTERVALO Distancia existente entre el valor máximo y el más bajo en un conjunto de datos.

INTERVALO DE INTERCUARTIL Diferencia entre los valores del primer y tercer cuartiles; esta diferencia indica el intervalo de la mitad intermedia del conjunto de datos.

INTERVALO DE INTERFRACTIL Medida de dispersión entre dos fractiles en una distribución; es decir, la diferencia que hay entre los valores de dos fractiles.

MEDIDA DE DISTANCIA Medida de dispersión en términos de la diferencia que existe entre dos valores en el conjunto de datos.

PERCENTILES Fractiles que dividen los datos en 100 partes iguales.

PUNTUACION ESTANDAR Hecho de expresar una observación en función de las unidades de desviación estándar por encima o por debajo de la media; es decir, la transformación de una observación restando la media y dividiendo entre la desviación estándar.

TEOREMA DE CHEBYSHEV Cualquiera que sea la forma de una distribución, por lo menos 75% de los valores de la población caerán dentro de 2 desviaciones estándar de la media y un mínimo de 89% caerá dentro de 3 desviaciones estándar.

VARIANCIA Medida del cuadrado de la distancia promedio entre la media y cada elemento de la población.

4-7 ECUACIONES UTILIZADAS EN EL CAPITULO

[4-1] $$\text{Intervalo} = \text{Valor de la observación más alta} - \text{Valor de la observación más baja}$$ *p. 122*

El *intervalo* es la diferencia que hay entre los valores máximo y mínimo en una distribución de frecuencia.

[4-2] $$\text{Intervalo del intercuartil} = Q_3 - Q_1$$ *p. 124*

El *intervalo del intercuartil* mide aproximadamente cuánto hemos de apartarnos de la media en ambos lados antes de poder incluir una mitad de los valores del conjunto de datos. Para calcular el intervalo dividimos los datos en cuatro partes iguales. Los *cuartiles* (Q) son los valores más altos en las cuatro partes. El intervalo de intercuartil es la diferencia existente entre los valores del primer y tercer cuartiles (Q_1 y Q_3).

[4-3] $$\text{Desviación del cuartil} = \frac{Q_3 - Q_1}{2}$$ *p. 125*

La *desviación del cuartil* mide la dispersión de una cuarta parte de los datos en una distribución. Es igual a una mitad del intervalo del intercuartil.

[4-4] $$\text{Desviación absoluta promedio} = \frac{\Sigma|x-\mu|}{N}$$ *p. 128*

Esta fórmula nos permite calcular la *desviación absoluta promedio de una población*. Esta medida de variabilidad se ocupa del valor absoluto de la diferencia entre cada elemento en el conjunto de datos y la media; no es tan útil para un cálculo ulterior como la variancia, que eleva al cuadrado cada distancia.

Aquí x representa el elemento u observación, μ la media de la población, N el número de elementos de la población y $\Sigma|x-\mu|$ la suma de todos los valores de $|x-\mu|$.

[4-5] $$\text{Desviación absoluta promedio} = \frac{\Sigma|x-\bar{x}|}{n}$$ *p. 129*

En una *muestra*, esta fórmula sirve para determinar la desviación absoluta promedio. A diferencia de la ecuación 4-4, se vale de la media muestral $\bar{x}$ y de los números de elementos en la muestra n.

[4-6] $$\sigma^2 = \frac{\Sigma(x-\mu)^2}{N} = \frac{\Sigma x^2}{N} - \mu^2$$ *p. 130*

Esta fórmula nos permite calcular la *variancia de la población*, o sea una medida del *cuadrado* de la distancia promedio que hay entre la media y cada elemento de la población. La expresión de en medio, $\frac{\Sigma(x-\mu)^2}{N}$, es la definición de σ^2. La última expresión, $\frac{\Sigma x^2}{N} - \mu^2$, es matemáticamente equivalente a la definición, pero a menudo resulta mucho más fácil de utilizar, pues con ella no es necesario calcular las desviaciones respecto de la media.

[4-7] $$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\Sigma(x-\mu)^2}{N}} = \sqrt{\frac{\Sigma x^2}{N} - \mu^2}$$ *p. 131*

La desviación estándar de la población, σ, es la raíz cuadrada de la variancia de la población. Es un parámetro más útil que la variancia ya que se expresa en las mismas unidades que los datos (mientras que las unidades de la variancia son los cuadrados de las unidades de los datos). Adviértase que la desviación estándar siempre es la raíz cuadrada *positiva* de la variancia.

[4-8] $$\text{Puntuación estándar de la población} = \frac{x-\mu}{\sigma}$$ *p. 133*

La *puntuación estándar* de una observación es el número de desviaciones estándar que la observación se encuentra por encima o por debajo de la media de la distribución. Esta puntuación nos permite hacer comparaciones entre los elementos de la distribución que difieren en el orden de magnitud o en las unidades empleadas. Aplicamos la ecuación 4-8 para obtener la puntuación estándar de un elemento de la *población*.

[4-9] $$\sigma^2 = \frac{\Sigma f(x-\mu)^2}{N} = \frac{\Sigma fx^2}{N} - \mu^2$$ *p. 134*

Esta fórmula en cualquiera de sus dos formas nos permite calcular la *variancia* de los datos ya *agrupados* en una distribución de frecuencia. Aquí, f, representa la frecuencia de clase y x representa la marca de clase.

[4-10] $$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\Sigma f(x-\mu)^2}{N}} = \sqrt{\frac{\Sigma fx^2}{N} - \mu^2}$$ *p. 134*

Se saca la raíz de la variancia y se obtiene así la *desviación estándar usando datos agrupados*.

[4-11] $$s^2 = \frac{\Sigma(x-\bar{x})^2}{n-1} = \frac{\Sigma x^2}{n-1} - \frac{n\bar{x}^2}{n-1}$$ *p. 135*

Para calcular la *variancia de la muestra*, se aplica la misma fórmula que en la ecuación 4-6, sustituyendo μ por $\bar{x}$ y N por $n-1$. El capítulo 8 contiene una explicación del por qué utilizamos $n-1$ en vez de n para obtener la variancia de la muestra.

[4-12] $$s = \sqrt{s^2} = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}} = \sqrt{\frac{\Sigma x^2}{n-1} - \frac{n\bar{x}^2}{n-1}}$$ *p. 135*

La *desviación estándar de la muestra* es la raíz cuadrada de la variancia de la muestra. Se parece a la ecuación 4-7, salvo que μ es reemplazada por la media muestral $\bar{x}$ y N se convierte en $n-1$.

[4-13] Puntuación estándar de la muestra $= \dfrac{x-\bar{x}}{s}$ *p. 138*

Esta ecuación se emplea para encontrar la puntuación estándar de un elemento de la *muestra*.

[4-14] Coeficiente de variación de la población $= \dfrac{\sigma}{\mu}(100)$ *p. 142*

El *coeficiente de variación* es una medida relativa de dispersión, la cual nos permite comparar dos distribuciones. Relaciona la desviación estándar y la media al expresar la primera como un porcentaje de la segunda.

4-8 EJERCICIOS DE REPASO

4-39 Una fábrica tiene un contrato con uno de sus clientes para suministrarle equipo de fresado para bombas. Una cláusula del contrato establece que el diámetro del equipo no debe exceder ciertos límites. A continuación se ofrecen los diámetros (en pulgadas) de una muestra de 20 equipos:

4.01	4.00	4.02	4.02	4.03	4.00	3.98	3.99	3.99	4.01
3.99	3.98	3.97	4.00	4.02	4.01	4.02	4.00	4.01	3.99

¿Qué puede decir Johnson a su cliente acerca de los diámetros de 95% del equipo que este último está recibiendo?

4-40 ¿Cómo reaccionaría usted ante la siguiente afirmación de un aficionado de fútbol americano: "Los Rockland Raiders promedian 3.6 yardas por acarreo en su juego terrestre. Dado que sólo necesitan 10 yardas para conseguir el primero y diez y tienen cuatro oportunidades para lograrlo, lo conseguirán con sólo recurrir a su juego por tierra."

4-41 ¿Cómo contestaría a la siguiente afirmación: "La variabilidad no es un factor importante, pues aun cuando el resultado sea menos seguro hay igual probabilidad de que caiga encima o debajo de la media. Por tanto, los resultados serán iguales en el promedio"?

4-42 A continuación se dan 3 secciones generales del presupuesto anual de defensa, cada una de las cuales recibió el mismo subsidio del Congreso:
a) Sueldos de funcionarios (total) b) Mantenimiento de aviones c) compras de víveres (total)
Teniendo en cuenta la distribución de los resultados posibles de los subsidios realmente destinados a estas áreas, correlacione cada sección de una de las curvas en la figura 4-1. Justifique su respuesta.

4-43 Una tienda de artículos deportivos almacena dos grados de cuerdas de pesca. He aquí los datos de las cuerdas:

	Fuerza media en la prueba (lbs)	Desviación estándar
Master	40	Se ignora el valor exacto, pero se estima que es muy grande
Super	30	Se ignora el valor exacto, pero se estima que es muy pequeño

Si usted va a pescar el pez azulado y plateado, que ha tenido un peso de 25 libras en esta temporada, ¿cuál cuerda de pesca probablemente atrape más pescados?

4-44 El vicepresidente de ventas de una compañía ha estado estudiando los registros referentes al desempeño de sus representantes de ventas. Se ha dado cuenta de que, en los 2 últimos años, el nivel promedio de ventas por representante ha permanecido inalterado, en tanto que la distribución del nivel se ha ampliado. Los niveles de ventas de vendedores a partir de ese periodo han presentado variaciones mucho más grandes respecto de la media que en cualquiera de los periodos de los dos años anteriores para los cuales cuenta el vicepresidente con registros. ¿Qué conclusiones pueden extraerse de estas observaciones?

4-45 Los automóviles nuevos que se vendieron durante diciembre en las 8 distribuidoras de la Ford, en un radio de 50 millas de Canton (Ohio) pueden describirse mediante el siguiente conjunto de datos:

124	234	187	197	188	456	345	190

a) Calcule el intervalo, la desviación absoluta promedio y la desviación estándar de estos datos.
b) ¿Cuál de las tres medidas que ha obtenido en (a) describe mejor la variabilidad de los datos?

4-46 Dos economistas están estudiando las fluctuaciones del precio del oro. Uno está analizando el periodo de 1968-1972. El otro examina el periodo comprendido entre 1975 y 1979. ¿Qué diferencia cabe esperar que presente la variabilidad de sus datos?

4-47 Una fábrica de botas tiene dos líneas de montaje en su planta. El gerente de fabricación quiere mejorar la uniformidad de la línea que ofrece la mayor variación. La línea número 1 fabrica un promedio mensual de 11,350 unidades, con una desviación estándar de 1,050. La línea número 2 fabrica un promedio mensual de 9,935, con una desviación estándar de 1,010. ¿Cuál de las dos líneas tiene la mayor dispersión relativa?

4-48 La estación de pesca y caza de Lake Wylie conserva registros de los peces atrapados en el lago y envía sus hallazgos al National Fish and Game Service. La captura en libras durante los últimos 20 días fue:

101	132	145	144	130	88	156	188	169	130
90	140	130	139	99	100	208	192	165	216

Calcule el intervalo, la variancia y la desviación estándar de los datos anteriores. ¿Es el intervalo una buena medida de variabilidad en este caso. ¿Por qué?

4-49 El dueño de una compañía grabadora, un gran minorista de discos, emplea 2 fórmulas diferentes para predecir sus ventas mensuales. La primera fórmula tiene un error promedio de 700 discos, con una desviación estándar de 35 discos. La segunda tiene un error promedio de 300 discos, con una desviación estándar de 16 discos. ¿Cuál de las dos fórmulas es relativamente menos exacta?

4-50 Con los siguientes datos de la población, calcule la desviación absoluta promedio (DAP), la variancia y la desviación estándar. ¿Qué le indican sus respuestas acerca del comportamiento de los costos de combustible de la calefacción?

COSTO PROMEDIO DEL COMBUSTIBLE DE CALEFACCION POR GALON EN OCHO ESTADOS

$1.65	$1.73	$1.66	$1.81	$1.79	$1.68	$1.74	$1.70

4-51 A continuación se da el número promedio de policías neoyorquinos de ambos sexos que diariamente trabajan en el distrito de Manhattan de las 8 a las 12 P.M.:

Lun. 2,950 Mar. 2,900 Miér. 2,900 Jue. 2,980 Vie. 3,285 Sáb. 3,430 Dom. 2,975

a) ¿Será la variancia o la desviación estándar una buena medida de la variabilidad de los datos precedentes?
b) ¿Qué aspecto del patrón de dotación de personal le hizo dar cierta respuesta a (a)?

4-52 Un psicólogo escribió un programa de computadora para simular la forma en que una persona llena un test estándar del coeficiente intelectual. Para probar el programa, introdujo en la computadora 15 formas diferentes de un test del coeficiente intelectual muy conocido y calculó el coeficiente en cada forma.

VALORES DEL CI

134	136	137	138	138
143	144	144	145	146
146	146	147	148	153

a) Calcule la media y la desviación estándar de las puntuaciones del coeficiente intelectual
b) Conforme al teorema de Chebyshev, ¿cuántos valores se encontrarán entre 132.44 y 153.56? ¿Y cuántos se hallan realmente en ese intervalo?

4-53 Cierto día, un departamento municipal de salubridad midió el peso de la basura en toneladas recogidas por sus 40 camiones de limpia. Los datos fueron dispuestos en el siguiente arreglo:

PESO DE LA BASURA (EN TONS)

16.2	15.8	15.5	15.3	15.0	14.9	14.9	14.8	14.7	14.6
14.6	14.5	14.5	14.4	14.3	14.0	13.9	13.9	13.8	13.7
13.5	13.2	13.0	12.9	12.7	12.4	12.2	12.0	12.0	11.9
11.8	11.5	11.4	11.1	11.0	10.9	10.9	10.0	9.5	9.1

Enumere los valores en cada decil. El 80% de los camiones recogieron menos de __________ toneladas.

4-54 La asistencia a los 10 últimos juegos del equipo de béisbol Baltimore Eagles fue la siguiente:

20,100	24,500	31,600	28,400	49,500
19,350	25,600	30,600	11,300	28,560

a) Calcule el intervalo, la desviación absoluta promedio, la variancia y la desviación estándar de los datos precedentes.
b) ¿Es una de sus respuestas en (a) una representación precisa de la variabilidad en los datos referentes a la asistencia?
c) ¿Qué otra medida de variabilidad podría ser una medida más adecuada?
d) Calcule el valor de la medida que usted sugiere en (c).

4-55 Una empresa de consultoría cuenta con los siguientes registros que indican el número de días en que sus 8 consultores se presentaron a trabajar:

212 220 230 210 228 229 231 219 221 222

a) Sin calcular el valor de ninguna de las medidas anteriores, ¿cuál de ellas cree usted que aportará más información sobre esta distribución: el intervalo, la desviación absoluta promedio, la desviación estándar?
b) Considerando la dificultad y el tiempo en el cálculo de cada una de las medidas que repasó en la parte (a), ¿cuál es la mejor a su juicio?
c) ¿Qué hará que usted cambie de opinión respecto a su elección?

4-9 AUTOEVALUACION

Las respuestas vienen al final del libro.

V F 1. La dispersión de un conjunto de datos nos permite conocer la confiabilidad de la medida de tendencia central.
V F 2. La desviación estándar es igual a la raíz cuadrada de la variancia.
V F 3. La diferencia entre la observación más alta y la más baja en un conjunto de datos recibe el nombre de intervalo de cuartil.
V F 4. La desviación del cuartil se basa sólo en dos valores tomados del conjunto de datos.
V F 5. La desviación estándar se mide en las mismas unidades que las observaciones en el conjunto de datos.
V F 6. El fractil es una localización en una distribución de frecuencia arriba o abajo de la cual se encuentra determinada proporción (o fracción) de los datos.
V F 7. La desviación absoluta promedio, a semejanza de la desviación estándar, tiene en cuenta todas las observaciones de un conjunto de datos.
V F 8. El coeficiente de variación es una medida absoluta de dispersión.
V F 9. La medida de dispersión de mayor uso entre los estadísticos es la desviación estándar.
V F 10. Una de las ventajas de las medidas de dispersión consiste en que cualquier estadístico que mida la variación absoluta mide también la variación relativa.
V F 11. Una desventaja del uso del intervalo para medir la dispersión consiste en que ignora la naturaleza de las variaciones entre la mayor parte de las observaciones.
V F 12. La variancia indica la distancia promedio que cualquier observación en los datos tiene con la media.
V F 13. Toda población tiene una variancia y ésta se representa mediante s^2.
V F 14. Según el teorema de Chebyshev, no más de 11% de las observaciones de una población pueden tener puntuaciones estándar mayores que tres o menores que -3.
V F 15. El intervalo del intercuartil es un ejemplo específico de un intervalo de interfractil.
V F 16. Es posible medir el intervalo en una distribución abierta.
V F 17. La desviación del cuartil mide el intervalo promedio del cuarto inferior de una distribución.
18. ¿Cuál de los siguientes enunciados es un ejemplo de una medida de distancia?

a) Intervalo c) desviación de intercuartil e) a y b pero no c
b) Intervalo de interfractil d) Todos los anteriores

19. ¿Cuál par de frases completa mejor esta proposición? Los fractiles que dividen los datos en ______________ partes iguales se llaman ______________.
a) 100 . . . deciles c) 10 . . . percentiles
b) 4 . . . cuartiles d) 16 . . . octiles

20. ¿Cómo se calcula una desviación del cuartil?
a) Se divide entre 4 el intervalo correspondiente del cuartil.
b) Se divide entre 2 el intervalo correspondiente del intercuartil.
c) Se multiplica por 4 el intervalo correspondiente del intercuartil.
d) Se multiplica por 2 el intervalo correspondiente del intercuartil.

21. ¿Por qué es necesario elevar al cuadrado las diferencias de la media cuando se calcula la variancia de la población?
a) Para que los valores extremos no afecten al cálculo.
b) Porque es posible que N sea muy pequeña.
c) Algunas de las diferencias serán positivas y otras serán negativas.
d) Ninguna de las razones anteriores.

22. Suponga que la población tiene $\mu = 100$, $\sigma = 10$. Si una observación en particular posee una puntuación estándar de 1, puede afirmarse que:
a) Su valor es 110.
b) Se encuentra entre 90 y 110, pero no puede determinarse su valor exacto.
c) Su valor es mayor que 110.
d) No puede determinarse nada si no se conoce N.

23. Suponga que una población tiene $\mu = 100$, $\sigma = 10$ y $N = 1{,}000$. Conforme al teorema de Chebyshev, ¿cuál de los siguientes enunciados NO es posible?
a) 150 valores son mayores que 130.
b) 930 valores se hallan entre 100 y 108.
c) 22 valores se encuentran entre 120 y 125.
d) 90 valores son menores que 70
e) Todos estos casos son posibles.

24. ¿Cuál de los siguientes casos es un ejemplo de una medida relativa de dispersión?
a) Desviación estándar c) Coeficiente de variación e) a y b pero no c
b) Variancia d) Todos los anteriores

25. ¿Cuál de los siguientes enunciados es verdadero?
a) La variancia puede calcularse para datos agrupados o no agrupados.
b) La desviación estándar puede calcularse para datos agrupados o no agrupados.
c) La desviación estándar puede calcularse para datos agrupados o no agrupados, pero la variancia puede calcularse sólo para datos no agrupados.
d) a y b pero no c.

26. Si tuviéramos que dividir la desviación estándar de una población entre la media de esa misma población y multiplicar este valor por 100, habríamos calculado:
a) La puntuación estándar de la población d) El coeficiente de variación de la población
b) La variancia de la población e) Ninguno de los anteriores
c) La desviación absoluta promedio

27. ¿En qué difiere el cálculo de una variancia de la muestra respecto al cálculo de la variancia de una población?
a) μ es sustituida por $\bar{x}$ d) a y c pero no b
b) N es sustituido por $n-1$ e) a y b pero no c
c) N es sustituido por n

28. El cuadrado de la variancia de una distribución es:
a) La desviación estándar c) El intervalo e) a y d
b) La media d) La desviación absoluta f) Ninguno de los anteriores

29. El teorema de Chebyshev establece que 99% de los valores se encontrarán a 3 desviaciones estándar positivas y negativas de la media para:
a) Distribuciones en forma de campana d) Todas las distribuciones
b) Distribuciones platicúrticas e) Ninguna distribución
c) Distribuciones de extremo izquierdo

30. En una distribución de frecuencia, la mediana es .5 ______________ porque la mitad del conjunto de datos es menor que ese valor o igual a él.

31. La diferencia entre los valores del primer y tercer cuartiles es el ______________ intervalo.

32. La medida del cuadrado de la distancia promedio entre la media y los elementos de la población es ______________. La raíz cuadrada positiva de este valor es ______________.

33. La expresión de la desviación estándar como porcentaje de la media es ______________.

34. El número de unidades de desviaciones estándar a que una observación se encuentra arriba o debajo de la media recibe el nombre de ______________.

35. Los fractiles que dividen los datos en 100 partes iguales se llaman ______________.

4-10 CASO CONCEPTUAL (Northern White Metals Company)

Al contar con una historia de tres años, detallada y ahora bien sintetizada, sobre la Northern White Metals Company, Dick estaba seguro de que su presentación tendría un éxito rotundo. Pero al repasar sus apuntes y cifras, lo invadió un repentino temor. "Esta presentación dista mucho de ser completa", pensó. "Algunos departamentos muestran muy altas tasas promedio de crecimiento de ventas, pero las cifras mensuales tienden a fluctuar ampliamente. Otros departamentos tienen una tasa menor, pero en un volumen de ventas más alto que al parecer fluctúa muy poco."

Situaciones similares se observaron en las ventas agrupadas por clase de clientes. Al examinar detenidamente ciertas agrupaciones de clientes, Dick había advertido la creciente importancia que las empresas de alta tecnología tenían para su compañía. Sin embargo, también en este caso el área de gran crecimiento parecía estar caracterizada por amplias variaciones en los datos observados. También la cobranza de las cuentas por cobrar en este grupo tendía a acumularse en exceso. La cobranza lenta siempre ocasionaba problemas, y éstos se agravaban a causa de la gran fluctuación en los precios primarios de los lingotes de aluminio que había habido en los últimos años.

"Desde luego", pensó Dick, "la estrategia que formula la compañía respecto a las clases de departamentos y clientes debe considerar la verdadera confiabilidad de estas medidas descriptivas que hemos obtenido. Antes que acabe de preparar esta presentación debo aprender a distinguir la uniformidad y confiabilidad de los datos acumulados." Mientras reflexionaba sobre lo que confiaba que sería la última fase de su presentación, sono el teléfono.

"Lennox", le espetó la voz ya muy conocida del presidente ejecutivo, "estamos esperando con interés tu presentación del lunes. Gran parte de la información que nos des servirá para la asignación del presupuesto de 1982. ¿Cómo va eso?"

"Estamos a punto de concluir, señor", replicó Dick. "Estoy examinando en estos momentos el volumen de ventas y ya analicé las cifras para hacernos una idea de las tendencias en varias áreas de nuestras operaciones. También estoy repasando el volumen de producción y las compras. Hice además algunos cálculos con objeto de facilitar la interpretación de un acervo bastante extenso de datos de apoyo." Luego añadió: "He observado algún comportamiento interesante en los patrones de ventas y del flujo de efectivo. . ."

"Eso está muy bien," contestó el presidente ejecutivo, interrumpiendo a Dick. "Creo que estás en la dirección correcta."

Dick colgó tranquilizado el teléfono. "Y ahora reanudemos nuestro trabajo."

Dick considera que su presentación, aunque de gran alcance, quedará incompleta si no se acompaña de un análisis ulterior. Se ha percatado de que ciertas cifras tienden a fluctuar mucho más que otras. Como no quiere hacer recomendaciones apresuradas ni erróneas, decide que convendría evaluar la variabilidad de los datos que va a aportar. ¿Cómo debe hacerlo? ¿Qué medidas le serán útiles? ¿Cómo serán interpretadas y utilizadas en una función de planeación?

4-11 EJERCICIO CON LA BASE DE DATOS

El miércoles por la mañana, después de otra sesión de perforación de tarjetas que se prolongó hasta muy noche, Frank y Laurel recibieron una llamada telefónica de Joe.

"Mañana a las 10 A.M. tenemos una reunión para tratar lo de las estrategias y en ella examinaremos además esta disminución de ventas. Fred presentará su propuesta, y yo quiero que ustedes dos hagan una recomendación. Preparen algo y, si tienen una buena sugerencia que ofrecer, después veremos los detalles."

"Dios mío," exclamó Frank al colgar el audífono. "Ni siquiera tenemos una idea clara. Y ésta puede ser nuestra única oportunidad de afianzarnos en nuestro puesto."

"Tranquilízate," contestó Laurel. "Ya hemos introducido los datos; ahora comenzamos a probar la variabilidad."

PROBLEMAS Y PREGUNTAS

1. ¿Cuál es la desviación absoluta promedio del tamaño medio de las órdenes cada año? (Use los datos del Capítulo 2.)
2. ¿Cuál es la variancia y la desviación estándar de cada año? (Use los datos brutos.) Con los datos agrupados del histograma que obtuvo en el capítulo 2, calcule la variancia y la desviación estándar correspondientes a 1985.
3. Aplique el teorema de Chebyshev para calcular el intervalo del tamaño de las órdenes que incluya por lo menos 89% de ellas.
4. Examine los histogramas trazados en el capítulo 2 y compárelos con los intervalos de Chebyshev calculados antes. ¿Con qué precisión establece el teorema de Chebyshev el intervalo en cada caso?
5. ¿En qué año hubo mayores probabilidades de que el tamaño de las órdenes individuales fuera de 700 o más?
6. ¿Cómo presentaría su análisis a Fred? ¿Tiene Fred razón al promover las ventas en las tiendas pequeñas exclusivamente? ¿Cómo concilia usted con su postura el hecho de que las ventas estén disminuyendo?
7. Cold River ha estado haciendo esfuerzos equivalentes de promoción tanto en las tiendas pequeñas como en las grandes. Supongamos que, en la reunión referente a las estrategias, se decida que la estrategia de Fred debe aceptarse. ¿Qué patrón mostrarán los datos de 1986? ¿Y en qué diferirán del patrón actual? ¿Apoyarán los datos la estrategia de Fred? ¿En qué se distinguirán los datos de 1986 de los de años anteriores en cuanto a la precisión con que reflejan el ambiente del mercado?

4-12 DIAGRAMA DE FLUJO