Métodos Cuantitativos

El autor del programa de Métodos Cuantitativos, David Targett, es profesor de Sistemas de Información en la Escuela de Administración de la Universidad de Bath. Anteriormente ocupó el cargo de profesor adjunto de Ciencias de Toma de Decisiones en la Escuela de Negocios de Londres. El profesor Targett tiene muchos años de experiencia en la formación de ejecutivos que desean incorporar herramientas numéricas a su acervo de destrezas gerenciales, y añadir así nuevos enfoques a sus metodologías de toma de decisiones. Su estilo se basa en la desmitificación de técnicas complejas y en la clara demostración de su relevancia, así como de sus limitaciones, en la vida real. Sus libros, entre los cuales se incluyen *Coping with Numbers ('Cómo Vérselas con Números')* y *The Economist Guide to Business Numeracy ('Guía de Técnicas Numéricas para Negocios publicado por The Economist')*, hacen énfasis en la comunicación por encima del rigor técnico, y han gozado de ventas a nivel mundial.

Ha escrito más de 50 estudios de casos prácticos, los cuales confirman la creciente integración de los métodos cuantitativos con otros aspectos de la gestión. Los casos abarcan una amplia variedad de sectores de la industria e ilustran la naturaleza cambiante de los métodos cuantitativos y las crecientes repercusiones que tienen en los individuos a cargo de la toma de decisiones en la era de la tecnología de la información. Asimismo, comprueban la vasta experiencia práctica acumulada por el Prof. Targett en organizaciones internacionales, tanto en el sector público como en el privado.

A uno de sus muchos artículos, un estudio sobre el suministro de información gerencial, le fue otorgado el Premio Pergamon en 1986.

Formó parte del equipo que diseñó el prestigioso programa de MBA a tiempo parcial de la Escuela de Negocios de Londres, del cual fue director entre los años 1985 y 1988. Durante este lapso, amplió el alcance internacional de la enseñanza, al dirigir los primeros grupos de estudio en Hong Kong, Singapur y en los Estados Unidos. Ha impartido clases en todos los programas principales que ofrece la Escuela de Negocios de Londres. Ha desarrollado y dirigido cursos sobre gerencia para un gran número de importantes empresas, entre ellas:

British Rail (Ferroviaria Británica)

Citicorp

Marks and Spencer (almacenes del Reino Unido)

Shell

Publicación QM-A1.2

ISBN 0 273 60932 7

Edinburgh Business School
Heriot-Watt University

Métodos Cuantitativos

Profesor David Targett

PEARSON Education

Edinburgh Gate, Harlow, Essex CM20 2JE, United Kingdom
Tel: +44 (0) 1279 623 623
Fax: +44 (0) 1279 431 059

Sitio web de Pearson Education: www.pearsoned.co.uk

Una empresa del grupo Pearson

Primera edición publicada en Gran Bretaña el 2003

ISBN 0 273 60932 7

British Library Cataloguing in Publication Data
Se puede obtener un catálogo CIP de este libro en la British Library.

Publicación QM-A1.2

CAPDM Ltd es responsable de la composición y gestión de los archivos SGML/XML. (www.capdm.com)
Impreso y encuadernado en Gran Bretaña.

Esta editorial tiene por política la utilización de papel procedente de bosques ecológicos.

Contenido

PARTE UNO

Presentación y trayectoria profesional

Módulo 1

Introducción a la Estadística: Algunos Usos Sencillos, Correctos e Incorrectos

Contenido

Lectura previa requerida: ninguna

Objetivos de Aprendizaje

Este módulo proporciona una visión general de la estadística. Presenta ideas y conceptos básicos a manera de generalidades, antes de tratarlos con más profundidad en módulos posteriores. Su objetivo es ofrecer una forma sencilla de adentrarse en el tema, para aquellos que no tienen conocimientos previos de estadística. Es una respuesta a la opinión cínica que afirma que para leer un texto sobre estadística se tiene que haberlo leído antes. Para quienes tengan conocimientos previos de estadística, el módulo les proporcionará un amplio marco de referencia para el estudio del tema.

1.1 Introducción

Los términos 'estadísticas' y 'estadística' hacen referencia, respectivamente, a una serie de números y a la ciencia dedicada al estudio de series de números. Bajo cualquier definición, este tema ha sido el blanco de bastante abuso, como se evidencia en la expresión 'mentiras, malditas mentiras y estadísticas'. Es posible que se deba a que, en buena medida, la gente no ha entendido que la estadística es similar a un lenguaje. De la misma manera en la que se puede dar mal uso al lenguaje hablado o escrito (por ejemplo, por los políticos y periodistas), el lenguaje numérico de la estadística también se puede utilizar de forma incorrecta (por los mismos políticos y periodistas). Culpar a la estadística por esto es tan poco sensato como sería culpar al idioma castellano cuando no se cumplen las promesas electorales.

No es necesario saber de estadística para hacer un mal uso deliberado de ella (otra expresión: 'los cálculos pueden mentir y los mentirosos pueden calcular'). No obstante, los malos usos a menudo pasan desapercibidos, ya que pocas personas parecen tener los conocimientos (y la seguridad en su uso) para manejar números. Comparativamente, hay muchas más personas que sí manejan bien las palabras. El número de personas con buenos conocimientos numéricos es menor que el de personas con buenos conocimientos lingüísticos. Sin embargo, lo que hace falta para detectar los malos usos de la estadística es el sentido común, con una pequeña cantidad de conocimientos técnicos.

Las dificultades se multiplican por las actitudes poco realistas de aquellos que verdaderamente tienen conocimientos de la estadística. Por ejemplo, si las memorias anuales de una empresa indican que el valor del inventario físico es de £34 236 417 (o aun si se indica £34 236 000), la afirmación se torna verosímil debido a la precisión de las cifras. Si presenciáramos el cálculo de las cifras realizado por los contadores, uno podría pensar que el método de recopilación de datos no justifica semejante grado de precisión. Una investigación de mercado que indique que 9 de cada 10 perros prefieren la comida para perros Bonzo es también engañosa, pero de una manera mucho más evidente. En última instancia, la afirmación carece de sentido, como se ve al hacer las siguientes preguntas: '¿En vez de qué otra comida?', '¿En qué circunstancias?', '¿9 de qué grupo de 10 perros?'

Tales ejemplos y muchos, muchos más de mayor o menor sutileza, han dado una mala reputación a la estadística, lo que con frecuencia se emplea como excusa para dejarla a un lado. Desafortunadamente, es imposible prescindir de la estadística en el mundo de los negocios. La toma de decisiones se basa en la información y ésta aparece a menudo en forma numérica. Para tomar buenas decisiones, es necesario organizar y entender los números. Esta es la razón de ser de la estadística y el porqué es tan importante tener algunos conocimientos acerca de ella.

La estadística puede dividirse en dos partes. La primera parte puede denominarse **estadística descriptiva**. A grandes rasgos, esta primera parte se ocupa del problema de ordenar una gran cantidad de datos, con el fin de hacer evidente sus principales características. Se trata de transformar números en información real y útil. Aquí se encuentran conceptos sencillos, como por ejemplo, la organización y preparación de datos a fin de poder apreciar sus patrones y también el resumen de los datos para facilitar su manejo y su comunicación a otros. Asimismo, incluye un aspecto que tiene mucha importancia en la actualidad: el tratamiento computarizado de las estadísticas empresariales, en la estructura en que la suministran los sistemas para la gestión de información y los sistemas de apoyo a la toma de decisiones.

La segunda parte puede denominarse a grandes rasgos como *estadística matemática o inferencia estadística*. Esta parte se ocupa del problema de cómo analizar una pequeña cantidad de información recopilada (denominada la **muestra**), a fin de inferir conclusiones generales acerca de la cantidad total de elementos similares que existen en el mundo (denominada la **población**). Un ejemplo son las encuestas de opinión, las cuales usan la inferencia estadística para hacer afirmaciones acerca del total del electorado de un país, con base en el resultado de apenas unos cientos de entrevistas.

Ambas partes de la estadística son vulnerables a los malos usos. Sin embargo, con unos pocos conocimientos y una gran cantidad de sentido común, se pueden identificar los errores y usar los procedimientos correctos. En este módulo se presentan los conceptos básicos de la estadística. Más adelante se indicarán algunos malos usos de la estadística y cómo hacerles frente.

La primera idea básica a la que hay que prestar atención es el concepto de probabilidad, fundamental para el trabajo estadístico. La estadística trabaja con aproximaciones y 'mejores estimaciones', dada la imprecisión y el carácter incompleto de la mayoría de los datos utilizados. Es poco usual que se hagan afirmaciones y que se extraigan conclusiones con certeza. La probabilidad es una manera de cuantificar qué tan sólida es la convicción que se tiene acerca de la información obtenida y de las conclusiones extraídas.

1.2 Probabilidad

Todos los acontecimientos futuros son inciertos hasta cierto punto. Es probable que el gobierno actual del Reino Unido esté aún en el poder el próximo año (suponiendo que dicho año no sea un año electoral), pero dista mucho de ser una certeza, mientras que el hecho de que un gobierno comunista siga en el poder el año próximo es altamente improbable, pero no imposible. La teoría de la probabilidad hace posible medir la incertidumbre relativa de algún suceso, mediante la medición de su probabilidad en una escala.

La escala se muestra en la Figura 1.1. En un extremo están los sucesos imposibles (por ejemplo, que una persona pueda cruzar el Océano Atlántico a nado). Éstos tienen una probabilidad de cero. En el extremo opuesto están los sucesos de total certidumbre (que algún día moriremos). Tienen una probabilidad

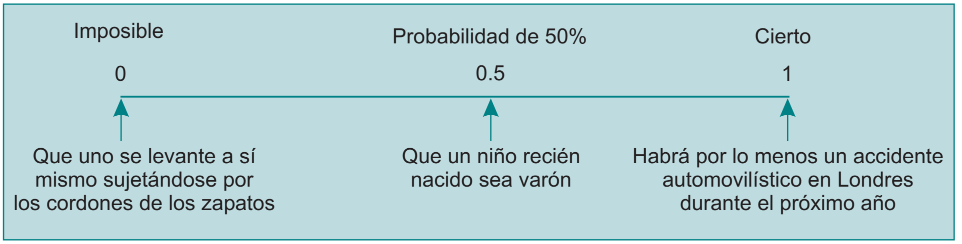

Figura 1.1 **Escala de probabilidad**

de uno. Entre ellos se sitúan todos los demás acontecimientos, que no son ni seguros ni imposibles y su posición viene dada en función de su grado de factibilidad. Por ejemplo, la probabilidad de que, al lanzar al aire una moneda, ésta caiga cara arriba, es de $\frac{1}{2}$; la probabilidad de que un determinado boleto sea el ganador de un sorteo en el que hay 100 boletos es de 0.01.

De forma abreviada, la expresión 'la probabilidad de un suceso A es 0.6' se representa de la siguiente manera:

$$P(\text{A}) = 0.6$$

1.2.1 Medición de la Probabilidad

Existen tres métodos para calcular la probabilidad. Para determinados sucesos, es posible que sólo uno de los tres métodos pueda ser el adecuado. Sin embargo, sí ofrecen formas conceptualmente distintas de entender la probabilidad. Esto quedará claro a medida que se describan los distintos métodos.

(a) **Enfoque a priori.** En este método, la probabilidad de un acontecimiento se calcula mediante un proceso de lógica. No es necesario realizar ningún experimento ni ponderación. Las probabilidades relativas a monedas, dados, juegos de naipes, etc. se pueden situar dentro de esta categoría. Por ejemplo, la probabilidad de que una moneda caiga de 'cara' varias veces seguidas se puede calcular teniendo en cuenta que la moneda tiene dos lados y que ambos lados tienen la misma oportunidad de quedar hacia arriba (los rebuscados deben tener en cuenta lo siguiente: se da por sentado que no quedará de canto). Como la moneda debe caer con uno de sus lados hacia arriba, los dos sucesos deben repartir de forma equitativa la probabilidad total de 1.0. Por lo tanto:

$$P(\text{Caras}) = 0.5$$
$$P(\text{Cruces}) = 0.5$$

(b) **Enfoque de la 'frecuencia relativa'.** Cuando el suceso se ha repetido o se puede repetir un gran número de veces, su probabilidad se puede medir a partir de la fórmula

$$P(\text{suceso}) = \frac{\text{Número de veces que se ha producido el suceso}}{\text{Número de repeticiones}}$$

Por ejemplo, para estimar la probabilidad de lluvia en un determinado día de septiembre en Londres, se pueden revisar los registros del clima de los últimos 10 años y encontrar que llovió en un total de 57 días. Por lo tanto:

$$\begin{aligned} P(\text{lluvia}) &= \frac{\text{Número de días de lluvia registrados}}{\text{Número total de días } (= 10 \times 30)} \\ &= \frac{57}{300} \\ &= 0.19 \end{aligned}$$

(c) **Enfoque subjetivo.** Un determinado sector de los profesionales de la estadística (los bayesianos), argumentaría que el grado de creencia que una persona tenga acerca de un suceso en particular se puede expresar como una probabilidad. Los bayesianos argumentan que en determinadas circunstancias, la evaluación subjetiva y personal de una probabilidad puede y debe utilizarse. La visión tradicional, sostenida por el sector de los profesionales de la estadística clásica, es que sólo son aceptables las evaluaciones objetivas de probabilidad. Las áreas y las técnicas que se sirven de las probabilidades subjetivas serán descritas más adelante. En este punto es importante conocer que las probabilidades pueden evaluarse de forma subjetiva, pero que existe un debate entre los profesionales acerca de la validez de dicho enfoque. Como ejemplo del enfoque subjetivo, consideremos como el suceso en cuestión la unidad política de Europa para el año 2010. No hay manera de utilizar alguno de los primeros dos enfoques para calcular la probabilidad de este suceso. Sin embargo, una persona puede expresar su opinión personal acerca de la probabilidad de este suceso al compararlo con un suceso cuya probabilidad se conoce, por ejemplo: ¿es más o menos probable que obtener 'cara' al lanzar una moneda? Tras un prolongado proceso de comparación y comprobación, el resultado podría ser el siguiente:

$$P(\text{unidad política de Europa para el año 2010}) = 0.10$$

El proceso de evaluar en forma precisa una probabilidad subjetiva es un campo de estudios aparte y no se debe considerar como meras especulaciones.

En este módulo se ha dado una sencilla y poco rigurosa introducción a los tres métodos para determinar probabilidades.. Sin importar el método, una vez calculadas las probabilidades se manejan exactamente del mismo modo.

Ejemplos

1 ¿Cuál es la probabilidad de sacar un seis al tirar un dado? Con un enfoque a priori, existen seis posibles resultados: 1, 2, 3, 4, 5 ó 6. Todos los resultados son igualmente probables, por lo tanto:

$$P(\text{sacar un 6}) = \frac{1}{6}$$

2 ¿Cuál es la probabilidad de que se termine la construcción de un segundo Eurotúnel para el tránsito de automóviles antes del año 2025?

El enfoque subjetivo es el único posible, ya que el razonamiento lógico por sí solo no puede conducirnos a una respuesta y no existen observaciones previas. Mi evaluación es muy pequeña, alrededor del 0.02.

3 ¿Cómo se calcularía la probabilidad de que una moneda trucada cayera con la cara hacia arriba?

Sería posible usar un enfoque a priori, si alguien tuviera información sobre el comportamiento aerodinámico de la moneda. Un método más realista consistiría en realizar varios lanzamientos de prueba y contar el número de veces que cae cara arriba:

$$P(\text{obtener cara}) = \frac{\text{Número de caras obtenidas}}{\text{Número de repeticiones}}$$

4 ¿Cuál es la probabilidad de sacar un as al tomar un naipe de una baraja?

Usemos el método a priori. Existen 52 resultados posibles (uno por cada naipe) y la probabilidad de tomar uno en particular, digamos el as de diamantes, debe ser: $\frac{1}{52}$. Hay cuatro ases en una baraja, por lo tanto:

$$P(\text{sacar un as}) = \frac{4}{52} = \frac{1}{13}$$

1.3 Distribuciones Estadísticas Discretas

La probabilidad hace posible el estudio de otro elemento esencial en el trabajo estadístico: la distribución estadística. Puede considerarse como uno de los primeros pasos de la estadística descriptiva, o como una piedra angular de la estadística matemática o de inferencias. Se desarrollará en primer lugar como una técnica descriptiva. Supongamos que existe una recopilación de datos que inicialmente podría ser similar a Figura 1.2.

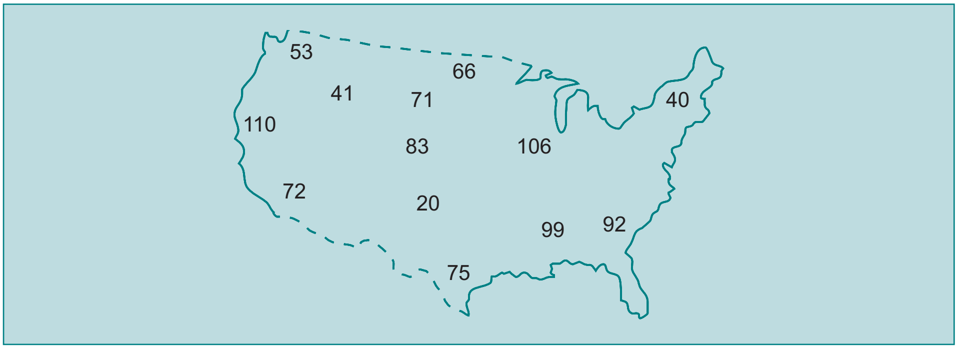

Figura 1.2 **Datos de ventas en EE. UU.**

Todas las cifras indican medidas de una variable. Se entiende por variable precisamente lo que implica el término. Es cualquier cosa que se pueda medir y cuya medición varía al realizar varias observaciones de la misma. La variable puede ser el número de delitos graves cometidos en cada departamento de

Francia, o la altura de todos los hombres de 20 años de Suecia. La Figura 1.2 muestra las ventas anuales (en miles) de una marca de maíz enlatado en diversas regiones de venta en los EE. UU. A los números se les denomina **observaciones** o **datos**.

No se ve sino desorden. Por supuesto, el desorden puede adoptar diferentes formas. El primer encuentro con un conjunto de datos puede ser sólo un montón de polvorientas notas de producción, o un archivo de facturas escritas a mano, pero siempre lo más probable es que estén desordenadas. Un primer intento de clasificación consistiría en ordenar los números, como la Tabla 1.1.

Tabla 1.1 Columnas de números

.	52	59	66
.	54	60	66
41	55	60	.
43	56	60	.
45	57	61	.
46	57	62	.
48	58	62	
49	58	63	
49	58	65	
50	59	65	

La Tabla 1.1 es una serie ordenada. Ahora se ve más ordenado, pero no es posible aún tener una idea acerca de los datos (por ejemplo, el promedio). El siguiente paso es hacer una clasificación de los datos y luego ordenar las clases. Por clasificar se entiende el agrupamiento de los números en intervalos o bandas (por ejemplo: 50–54), para facilitar su tratamiento. Cada clase tiene un valor de frecuencia, es decir, el número de datos que se sitúan en dicha clase. A esto se le denomina **tabla de frecuencias** y se muestra en la Tabla 1.2. Siete datos fueron mayores o iguales a 40, pero menores a 50; 12 eran mayores o iguales a 50, pero menores a 60; etc. En total había 100 datos.

Tabla 1.2 Una tabla de frecuencias

Banda	*Frecuencias*
$40 \leq x < 50$	7
$50 \leq x < 60$	12
$60 \leq x < 70$	22
$70 \leq x < 80$	27
$80 \leq x < 90$	19
$90 \leq x < 100$	10
$100 \leq x < 110$	3
Total de frecuencias	100

Nota: $\leq$ significa 'menor o igual a'; $<$ significa 'menor que'.

Ahora es mucho más fácil obtener una idea general de la información que aportan los datos. Por ejemplo, en este caso la mayoría de los números se encuentran entre 60 y 90, con extremos de 40 y 110. Por supuesto, es probable que en algún momento exista la necesidad de hacer cálculos detallados con estas cifras, a fin de obtener información específica. Pero en este momento, el objetivo se centra en obtener una idea general acerca de los datos en el menor tiempo posible. Otra presentación, de mayor impacto visual y que ayuda a lograr dicho objetivo, es el denominado **histograma de frecuencias**.

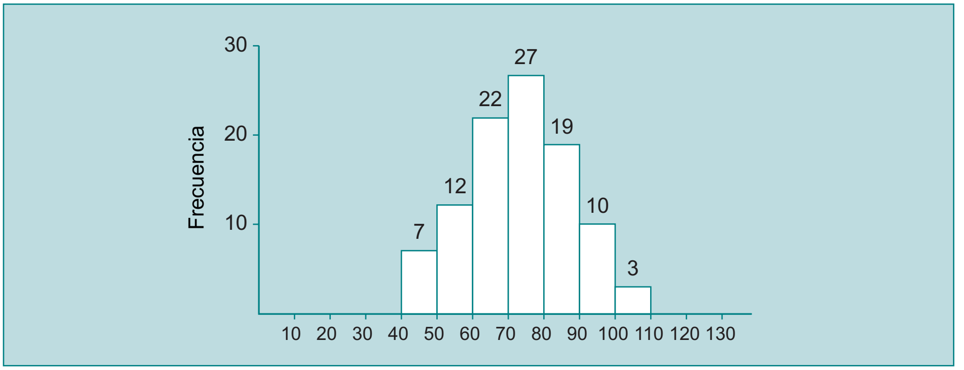

Figura 1.3 **Un histograma de frecuencias**

El paso de la Tabla 1.2 a la Figura 1.3 es sencillo y obvio. No obstante, el histograma de frecuencias permite captar de forma inmediata cómo son los datos. Las cifras están distribuidas de forma simétrica en un intervalo de valores que abarca desde 40 hasta un poco por encima de 100 y la mayoría se encuentra en el área central.

Para efectos descriptivos, el histograma de frecuencias funciona a la perfección y no es necesario refinarlo más. Si, por otra parte, hubiera objetivos analíticos, el histograma de la Figura 1.3 se transformaría en una distribución estadística. Para ser muy precisos, todas las formas de presentación de datos tienen que ver con la distribución estadística, pero lo que se procura es encontrar la versión más fácil de manejar y de mayor aceptación general.

Para llevar a cabo dicha transformación, hay que tomar en cuenta en primer lugar la vinculación entre frecuencias y probabilidades, mediante el enfoque de la 'frecuencia relativa' del cálculo de probabilidades. La probabilidad de que una medida seleccionada al azar esté situada en una clase determinada se puede calcular de la siguiente manera:

$$P(\text{número esté situado dentro de la banda } x) = \frac{\text{Banda de frecuencias } x}{\text{Total de frecuencias}}$$
$$= \frac{\text{Altura de la columna } x}{\text{Total de frecuencias}}$$

por ejemplo,

$$P(40 \leq x < 50) = \frac{7}{100} = 0.07$$

El histograma de frecuencias se puede convertir en un **histogramas de probabilidades**, si escribimos las unidades del eje vertical como probabilidades (tal como se calcularon antes) en lugar de frecuencias. La forma del histograma permanecería inalterada. Cuando el histograma se encuentra en el formato de probabilidad, usualmente se le denomina distribución (en este caso, una distribución discreta). Una variable es discreta si tiene limitaciones en los valores que puede adoptar. Por ejemplo, cuando los datos están restringidos a bandas o intervalos (como en el caso anterior) la variable es discreta. Cuando el valor de una variable está restringido a números enteros (una variable entera) también es discreta.

El histograma de probabilidades facilita el manejo de las probabilidades asociadas con uniones de clases. Por ejemplo, si las probabilidades de dos de las clases son:

$$P(50 \leq x < 60) = 0.12$$
$$P(60 \leq x < 70) = 0.22$$

entonces:

$$\begin{aligned} P(50 \leq x < 70) &= 0.12 + 0.22 \\ &= 0.34 \end{aligned}$$

Esto se cumple ya sea que se trabaje con probabilidades o con las frecuencias de las que se derivan.

Ejemplos

Partiendo de los datos de la Figura 1.3, ¿cuáles son las probabilidades de que:

1 $P(80 \leq x < 100)$?

2 $P(x < 70)$?

3 $P(60 \leq x < 100)$?

Respuestas

1 $$\begin{aligned} P(80 \leq x < 100) &= P(80 \leq x < 90) + P(90 \leq x < 100) \\ &= \frac{19}{100} + \frac{10}{100} \\ &= 0.19 + 0.10 \\ &= 0.29 \end{aligned}$$

2 $$\begin{aligned} P(x < 70) &= P(x \leq 50) + P(50 \leq x < 60) + P(60 \leq x < 70) \\ &= 0.07 + 0.12 + 0.22 \\ &= 0.41 \end{aligned}$$

3 $$\begin{aligned} P(60 \leq x < 100) &= 0.22 + 0.27 + 0.19 + 0.10 \\ &= 0.78 \end{aligned}$$

1.4 Distribuciones Estadísticas Continuas

Para resumir el progreso hasta el momento: existe un histograma de probabilidades de una variable, del cual se puede determinar la probabilidad de que alguna de las medidas de la variable se sitúe dentro de una de las clases o bandas del histograma. A este tipo de distribución se le denomina **distribución discreta**. Se trata de una distribución porque la variable se distribuye dentro de un intervalo de valores y es discreta porque los valores que adoptan las variables son escalonados, en lugar de ser variaciones consecutivas.

Una variable **continua** no está limitada en los valores que puede adoptar. Puede equivaler a números enteros y también a cualquier valor entre ellos. No agrupa los datos en clases, sino que distingue entre números tales como 41.73241 y 41.73242. La distribución que forma una variable continua se llama **distribución continua**. Se puede considerar como una extensión de la distribución discreta. El proceso de extensión es el siguiente. (Se muestra el proceso para ilustrar el vínculo entre las distribuciones discretas y las continuas. No se trata de un procedimiento que se necesita llevar a cabo en la práctica).

Una distribución discreta como la de la Figura 1.3 se repite en la Figura 1.4(c). El ancho de las columnas se reduce de forma progresiva. En (b) el ancho de las columnas se ha dividido en dos; por ejemplo, la clase $50 \leq x < 60$ se ha dividido en dos clases $50 \leq x < 55$ y $55 \leq x < 60$. En (c) las clases se han vuelto a dividir. A medida que avanza el proceso, se agiliza la distribución hasta que en última instancia, se logra la distribución continua (d).

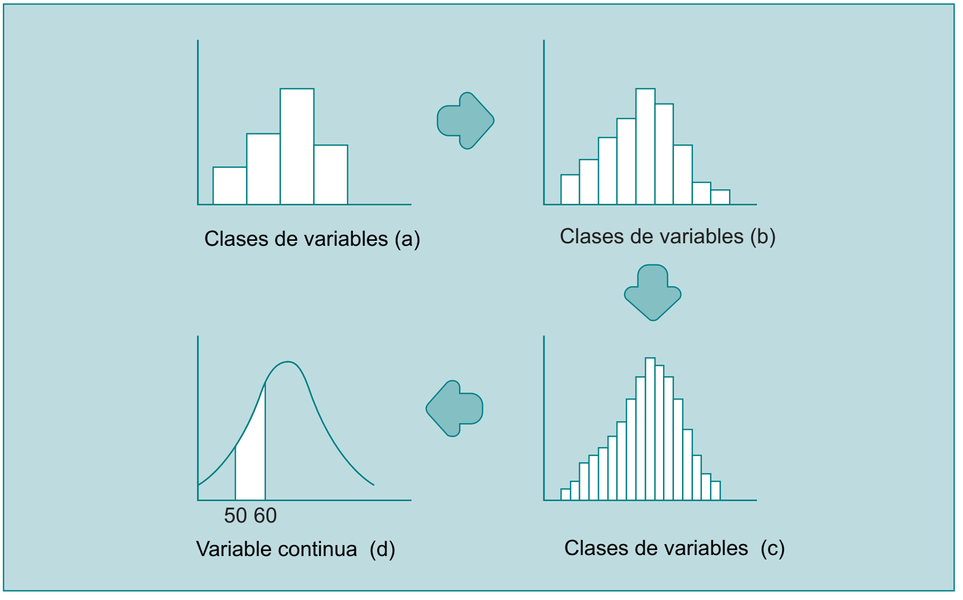

Figura 1.4 El paso de la función discreta a la continua

Ahora surge una dificultad en cuanto a la medición de probabilidades. En la distribución discreta, la Figura 1.4(a), las probabilidades asociadas a los diferentes valores de la variable son iguales a la altura de la columna. Si la altura de la columna sigue siendo igual a la probabilidad, el proceso (a) → (b) → (c) → (d) dará como resultado histogramas cada vez más planos. La Figura 1.4(d) será completamente horizontal, ya que la probabilidad asociada a los valores ahora separados, como 41.73241 y 41.73242, debe ser infinitesimalmente pequeña. El problema se solventa al medir las probabilidades en una distribución continua mediante *áreas*. Por ejemplo: $P(50 \leq x < 60)$ es el área bajo el segmento de la curva entre 50 y 60 y aparece sombreado en la Figura 1.4(d).

El argumento para utilizar las áreas es el siguiente: en la Figura 1.4(a) el ancho de las columnas es igual, por lo que la probabilidad puede medirse tanto por el área como por la altura. La Figura 1.5 proporciona un ejemplo de lo que ocurre en el paso de (a) a (b), al dividir las clases en dos. Se supone que los datos originales permiten que las probabilidades de las clases nuevas se puedan calcular a partir de las originales.

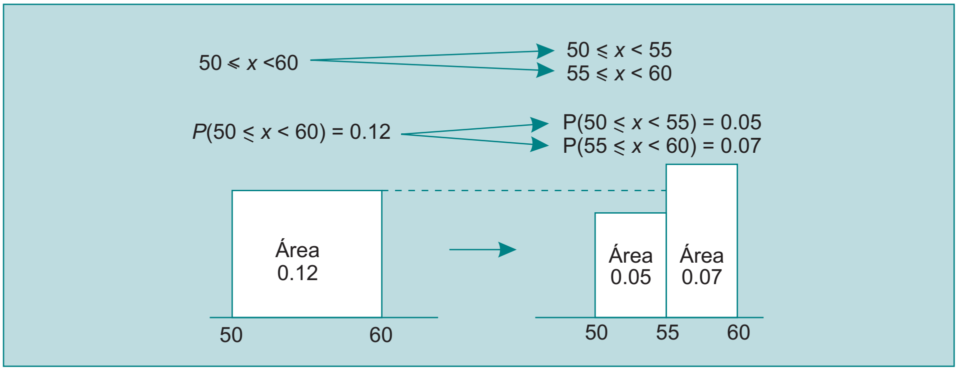

Figura 1.5 **Reducción del tamaño de las clases**

Al utilizar áreas para medir probabilidades, la altura de las columnas de las clases nuevas es aproximadamente la misma que la de la original. Las menores probabilidades de las clases nuevas se ven reflejadas en la división del ancho de las columnas, en lugar de por cambios de altura. A medida que el proceso de subdivisión continúa, desaparece la tendencia a que la distribución se haga más aplanada. De este modo, una distribución continua puede tener una forma definida, que puede interpretarse del mismo modo que la forma de una distribución discreta, pero la probabilidad se mide a partir de las áreas. Al igual que la altura de las columnas de una distribución discreta suman 1 (debido a que cada observación posee *ciertamente* algún valor), el área total de una distribución continua también es 1.

Las diferencias entre las distribuciones discretas y las continuas se resumen en la Tabla 1.3.

Tabla 1.3 Diferencias entre distribuciones discretas y continuas

Discreta	*Continua*
Variable limitada a determinados valores	Variable no limitada
Usualmente de forma escalonada	Usualmente de contorno suave
Las probabilidades son iguales a la altura de las columnas	Las probabilidades son iguales al área situada bajo la curva
La suma de las alturas de las columnas = 1	El área total = 1

Ejemplo

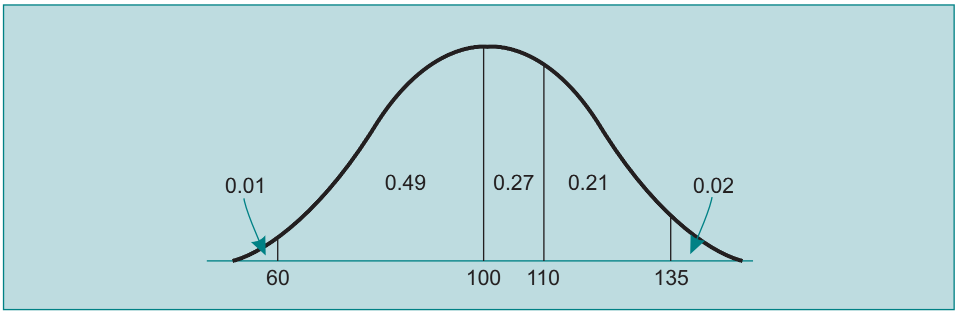

Figura 1.6 Se muestra el área situada debajo de cada parte de la curva. El área total es igual a 1.0

Si se utiliza la distribución continua de la Figura 1.6, ¿cuáles son las probabilidades de que un valor determinado de la variable se sitúe dentro de los siguientes intervalos?

1 $x \leq 60$?
2 $x \leq 100$?
3 $60 \leq x \lesssim 110$?
4 $x \geq 135$?
5 $x \geq 110$?

Respuestas

1 $P(x \leq 60) = 0.01$
2 $P(x \leq 100) = 0.01 + 0.49 = 0.5$
3 $P(60 \leq x \lesssim 110) = 0.49 + 0.27 = 0.76$
4 $P(x \geq 135) = 0.02$
5 $P(x \geq 110) = 0.21 + 0.02 = 0.23$

En la práctica, los problemas asociados al uso de las distribuciones continuas son, en primer lugar, el no poder recopilar nunca la cantidad suficiente de datos, ni medirlos con la precisión necesaria, como para establecer una distribución continua. En segundo lugar, si lo anterior fuera posible, sería difícil realizar mediciones precisas de las áreas bajo la curva. Su principal uso práctico está en considerar las distribuciones continuas como *distribuciones estándar*, un tema que se tratará en la sección siguiente.

1.5 Distribuciones Estándar

La distribución de las ventas en los EE. UU. que se muestra en la Figura 1.2, la Tabla 1.1, la Tabla 1.2 y en la Figura 1.3 es una distribución **observada**. Se recopilaron los datos, se hizo el histograma y se obtuvo dicha distribución. La distribución **estándar** tiene sus fundamentos en la teoría y no en las observaciones. Se trata de una distribución definida matemáticamente, a partir de una situación teórica. Las características de la situación se expresan matemáticamente y la situación resultante se construye de forma teórica. Cuando aparece una situación real que se asemeja a la teórica, se aplica la distribución estándar asociada.

Por ejemplo, un caso de distribución estándar, la **distribución normal**, es producto de la siguiente situación teórica. Se genera una variable que es el resultado de un proceso que debería dar valores constantes, pero no ocurre así, ya que dicho proceso está sujeto a una gran cantidad de perturbaciones menores. Como resultado, la variable se distribuye en torno al valor central (*véase* la Figura 1.7). Esta situación (valor central, numerosas perturbaciones menores) se puede expresar matemáticamente y la distribución resultante se puede predecir matemáticamente, es decir, se puede encontrar una fórmula matemática que la describa.

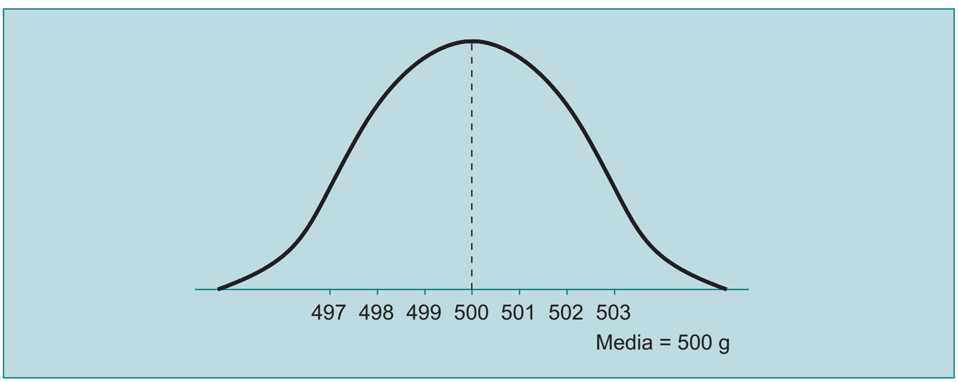

Figura 1.7 **Distribución normal del peso de las hogazas de pan**

Si una situación real parece ser similar a la teórica, se utiliza la distribución normal. Se puede entonces hacer cálculos de probabilidades, similares a los usados para las cifras de ventas en los EE. UU. Las áreas bajo los segmentos de la curva pueden calcularse a partir de su fórmula matemática o, lo que es más fácil, de las tablas de distribución normal. La distribución normal sería adecuada, por ejemplo, para la longitud de varillas cortadas a máquina. Todas las varillas deberían ser de la misma longitud, pero no lo son debido a variaciones introducidas por las vibraciones, las imprecisiones de la máquina, el operador, etc. Un análisis típico consistiría en calcular el porcentaje de la producción susceptible de quedar fuera de las tolerancias de diseño de las varillas.

La distribución normal se puede aplicar a muchas situaciones con características similares. Otras distribuciones estándar se refieren a situaciones con características diferentes. La aplicación de una distribución estándar es una aproximación. No es probable que la situación real se ajuste exactamente a la teórica, la que sirve de fundamento a la fórmula matemática. Sin embargo, esta desventaja se ve más que compensada, ya que el uso de la distribución estándar reduce la cantidad de datos que hay que recopilar. Comúnmente, las distribuciones observadas requieren la recopilación de una gran cantidad de datos. No sólo es necesario recopilar suficiente información para que la distribución tome forma, sino que los datos también se deben recopilar de manera individual para todas y cada una de las situaciones.

En resumen, el uso de una **distribución observada** supone que se han recopilado datos y que se han hecho los histogramas. El uso de una **distribución estándar** supone que la situación de la cual se obtienen los datos se asemeja bastante a una situación teórica, para la cual se ha desarrollado una distribución matemáticamente.

1.5.1 La Distribución Normal

A continuación estudiaremos en profundidad la distribución normal, una de las más comunes. La Figura 1.7 presenta una visión a grandes rasgos de su forma, para el caso del peso de hogazas de pan. Sus características principales son la simetría y la forma de campana, con una sola joroba (es decir, es **unimodal**), situada en el promedio o media aritmética de la variable.

Sin embargo, no todas las distribuciones normales son idénticas. De lo contrario, no sería posible que pudieran representar tanto el peso de hogazas de pan (con un valor promedio de 500 g y una variación de más o menos 10 g) como la altura de los hombres adultos (con un promedio de 1.75 metros y una variación de más o menos 0.40 m). Todas las curvas normales comparten una serie de propiedades comunes, como las mencionadas anteriormente; sus diferencias radican en que las poblaciones que describen tienen características diferentes. Dos factores, denominados **parámetros**, encarnan dichas características y son suficientes como para distinguir una curva normal de otra (y, a la inversa, para especificar totalmente una curva normal). Un parámetro se define como una medida que describe algún aspecto de una población.

El primer parámetro es el **promedio** o media de la distribución. A pesar de que el término 'promedio' no se ha definido formalmente con anterioridad, no es más que el concepto de uso cotidiano (por ejemplo: el promedio de dos y cuatro es tres). Dos distribuciones normales que difieran únicamente en este parámetro, tienen exactamente la misma forma, pero están situadas en distintos puntos a lo largo de una escala horizontal.

El segundo parámetro es la **desviación estándar**. Más adelante se dará su definición exacta. Mide la dispersión, o el alcance, de la variable. En otras palabras, algunas variables se agrupan estrechamente en torno al promedio (como las hogazas de pan). Dichas distribuciones presentan una desviación estándar baja y su forma es estrecha y alta. Las variables que están muy dispersas alrededor del promedio presentan una desviación estándar elevada y su distribución es baja y aplanada. La Figura 1.8 muestra ejemplos de distribuciones con desviaciones estándar altas y bajas: los salarios en un hospital tienen una gran dispersión, que va desde los de los empleados de limpieza hasta los de los asesores; los salarios del personal docente de una escuela tiene una dispersión mucho menor.

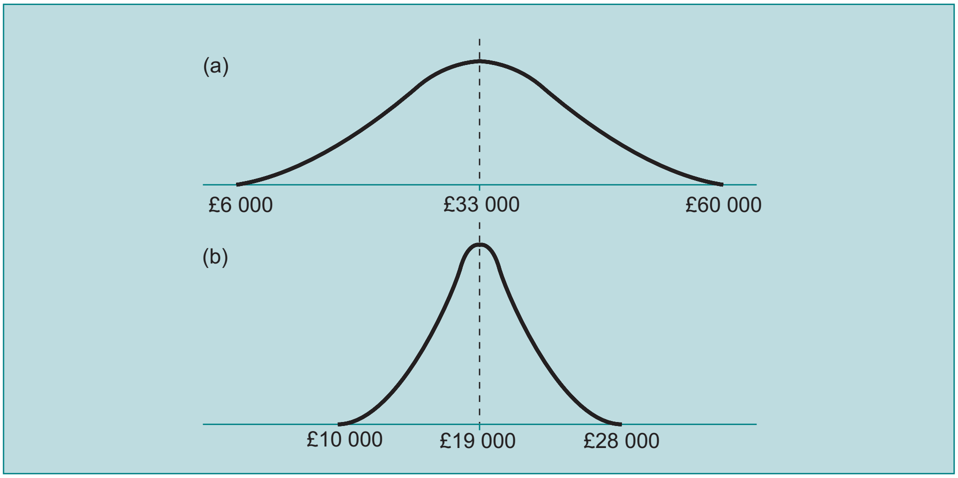

Figura 1.8 **Salarios: (a) hospital: desviación estándar alta; (b) escuela: desviación estándar baja**

Otra característica de una distribución normal está vinculada a la desviación estándar (*véase* la Figura 1.9). Los datos se refieren al peso de las hogazas de pan, con un peso promedio de 500 g y una desviación estándar de 2 g.

La propiedad de la distribución normal ilustrada en la Figura 1.9 se deriva de las matemáticas subyacentes, que están más allá del alcance de esta introducción. En cualquier caso, es más importante ser capaz de utilizar la distribución normal, que saber demostrar matemáticamente sus propiedades. La propiedad se aplica independientemente de si la distribución es aplanada y ancha o alta y estrecha, siempre que sea normal. Dada esa propiedad, es posible calcular las probabilidades de los sucesos. El ejemplo a continuación demuestra cómo se puede utilizar una distribución estándar (en este caso la normal) en el análisis estadístico.

Ejemplo

Una máquina está ajustada para producir componentes de acero de una longitud determinada. Se toma una muestra de 1 000 componentes y se miden sus longitudes. A partir de las mediciones, se estima que el promedio y la desviación estándar de todos los componentes producidos son de 2.96 cm y 0.025 cm respectivamente. ¿Cuáles son los límites entre los que se esperaría encontrar el 95% de todos los componentes producidos por la máquina?

Sigamos los pasos a continuación:

1. Vamos a suponer que la longitud de todos los componentes producidos sigue una distribución normal. Esto es razonable ya que esta situación es típica de las circunstancias en las que se ocurren las distribuciones normales.
2. Los parámetros de la distribución son los siguientes: media = 2.96 cm y desviación estándar = 0.025 cm. Por lo tanto, la distribución de la longitud de los componentes será tal y como se muestra en la Figura 1.10.

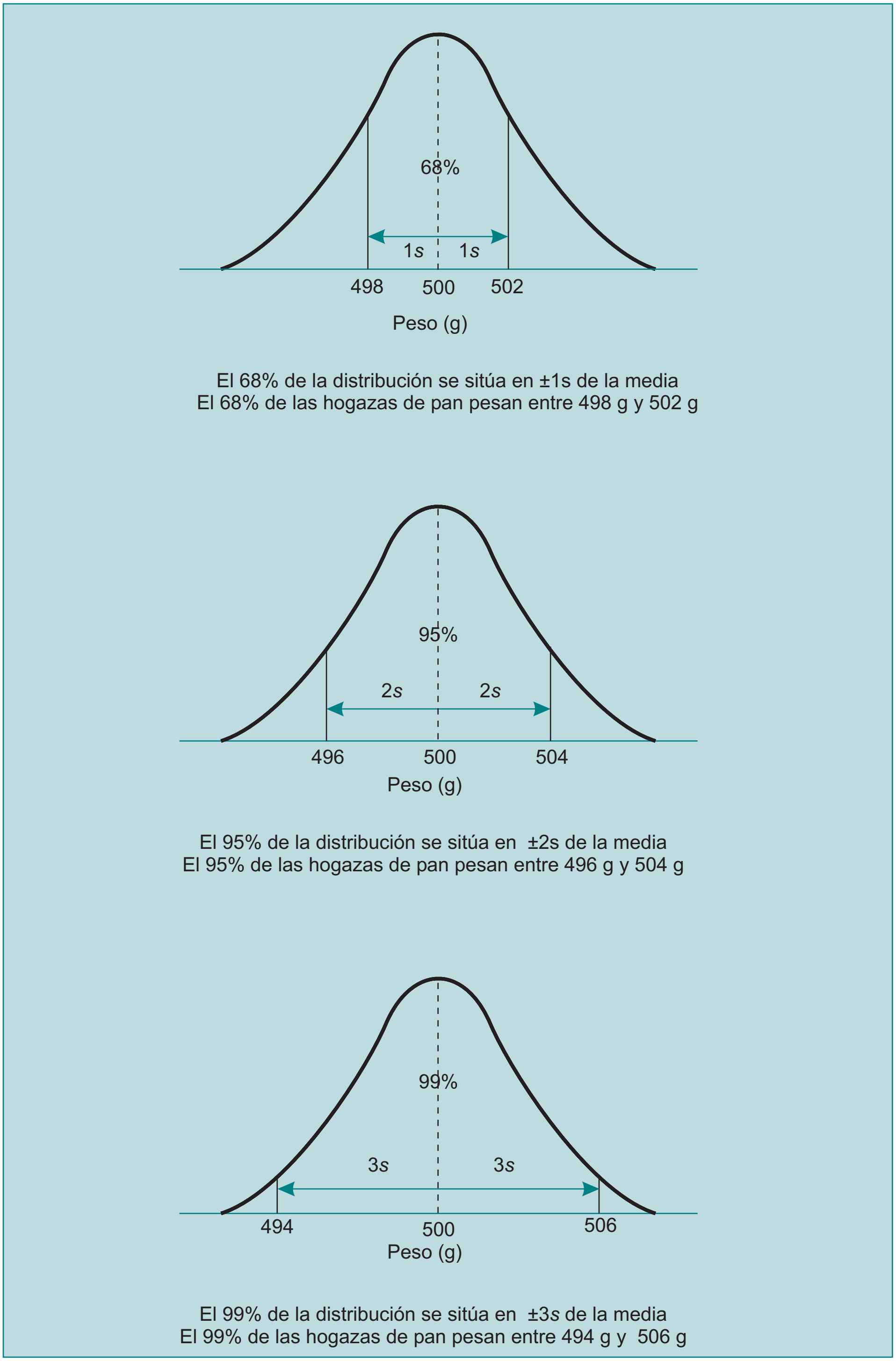

Figura 1.9 **Características de las desviaciones estándar en una distribución normal**

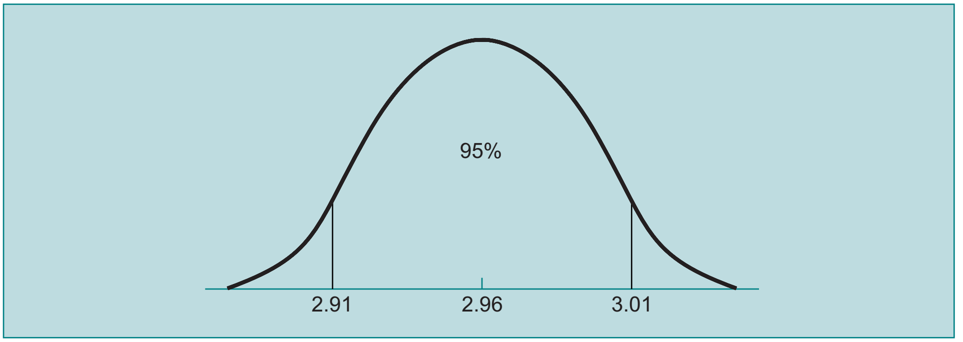

Figura 1.10 Distribución de la longitud de los componentes de acero

Existe una diferencia entre la distribución de *todos* los componentes producidos por la máquina (la distribución de la **población**) y la distribución de la longitud de los componentes en la muestra. La primera distribución es la relevante y se muestra en la Figura 1.10. La muestra se ha utilizado para estimar los parámetros.

3 A partir de las propiedades de la distribución normal enunciadas anteriormente, el 95% de la distribución de la población (y, por tanto, el 95% de todos los componentes producidos) se encontrará dentro de dos desviaciones estándar del promedio. Los límites son:

$$= 2.96 - (2 \times 0.025) \text{ y } 2.96 + (2 \times 0.025)$$
$$= 2.91 \text{ y } 3.01 \text{ cm}$$

De acuerdo con la estimación, el 95% del total de la producción se situará entre 2.91 cm y 3.01 cm.

1.6 Usos Incorrectos de la Estadística

La estadística se usa en forma incorrecta cuando la evidencia estadística se presenta de tal modo que fomenta conclusiones falsas. Las autoridades que supervisan la publicidad buscan proteger al público de la publicidad engañosa, pero los gerentes no disponen de una protección similar contra la información administrativa errónea. El presentar los datos de esa forma puede ser accidental o deliberado. En el último caso, el uso incorrecto de la estadística alcanza niveles casi de creación artística. Aún en esos casos, es posible detectar algunos usos incorrectos.

1.6.1 Definiciones

Es posible que no haya una definición exacta para las expresiones estadísticas y para las variables mismas. El usuario puede suponer que el que produce los datos trabaja con una definición diferente de la que realmente se usó. Al suponer una definición incorrecta, el usuario sacará conclusiones incorrectas La expresión estadística 'promedio' se presta a diversas interpretaciones. Una empresa de contabilidad anuncia en su folleto de reclutamiento que el salario promedio de los contadores calificados de la empresa es de £44 200. Visto así, un empleado potencial podría concluir que trabajar para esta empresa es económicamente atractivo. Una mirada más de cerca muestra que los contadores de la empresa y sus salarios son los siguientes:

3 accionistas	£86 000
8 contadores senior	£40 000
9 contadores junior	£34 000

El salario promedio podría ser:

$$\text{La media aritmética} = \frac{3 \times 86\,000 + 8 \times 40\,000 + 9 \times 34\,000}{20}$$
$$= £44\,200$$
$$\text{el valor 'intermedio'} = £40\,000$$
$$\text{o el valor más frecuente} = £34\,000$$

Se podría decir de manera legítima que cualquiera de estas cifras constituye el salario promedio. Indudablemente, la empresa ha elegido la que más se adapta a sus propósitos. Incluso si fuese cierto que se usa la definición estadística correcta, aún sería necesario preguntar cómo se define la variable (el salario). ¿Se incluye en el salario de los accionistas el reparto de la utilidad? ¿Están incluidas las bonificaciones en el salario de los contadores? ¿Están incluidos los beneficios extrasalariales (un automóvil, por ejemplo) en el salario de los contadores? Si se eliminan estos detalles, la situación podría ser la siguiente:

3 accionistas	£50 000
8 contadores senior	£37 000
9 contadores junior	£32 400

El salario medio es ahora de £36 880. De repente, la remuneración que ofrece la empresa ya no es tan atractiva.

1.6.2 Gráficos

Las imágenes visuales usadas en estadística buscan comunicar los datos con suma rapidez. Esta rapidez denota que las primeras impresiones son importantes. Si la primera impresión es incorrecta, no es muy factible que se corrija.

Las ilustraciones gráficas de datos son muchas, pero probablemente la más utilizada es el gráfico. Si la escala de un gráfico está disimulada, o si no aparece en absoluto, se pueden extraer conclusiones erróneas. La Figura 1.11 muestra las cifras de ventas de una compañía durante los últimos tres años. Podría parecer que la empresa ha tenido éxito.

Sin embargo, no se muestran las escalas. De hecho, las cifras de ventas han sido las siguientes:

1994	£11 250 000
1995	£11 400 000
1996	£11 650 000

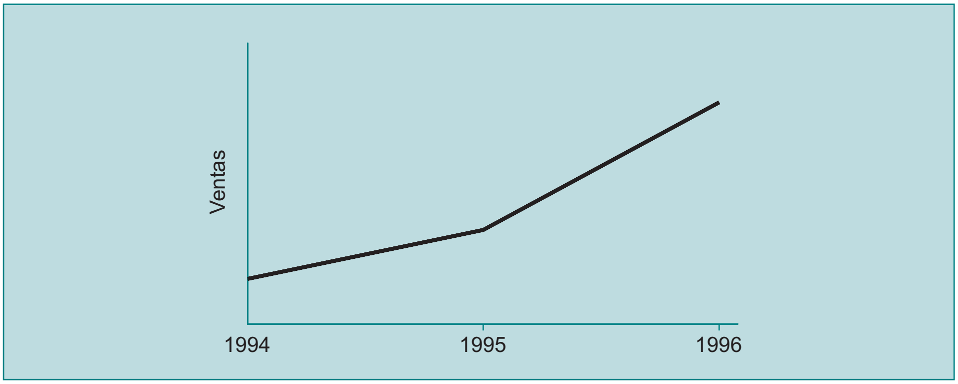

Figura 1.11 Cifras de ventas (sin escala)

En la Figura 1.12 hay un gráfico más informativo que sí muestra la escala. Las ventas apenas han aumentado. Si se considera la inflación, es probable que hayan disminuido en términos reales.

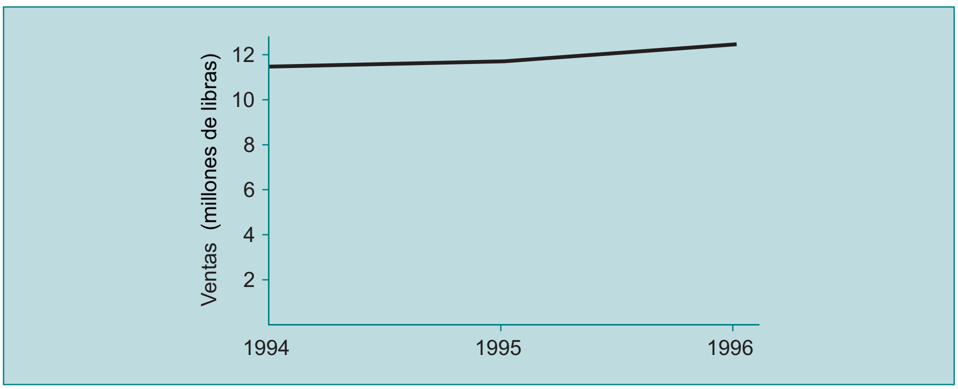

Figura 1.12 Cifras de ventas (con escala)

1.6.3 Sesgo o Error Sistemático

La mayor parte de los datos estadísticos se recopilan a partir de una muestra; en otras palabras, representan sólo una pequeña parte del total de los datos disponibles (la población). Las conclusiones extraídas de la muestra son extrapoladas al ámbito global de la población. La generalización puede ser válida sólo en la medida en que la muestra sea representativa. Si la muestra no es representativa, entonces se extraerán conclusiones erróneas. El sesgo en las muestras puede ocurrir de tres maneras.

En primer lugar, puede surgir a partir de la recopilación de datos. Consideremos el caso de un político de izquierda que afirma que el 80% de las cartas que recibe se muestran en contra de una medida del gobierno de derecha y concluye que una mayoría del electorado se opone al gobierno en ese asunto. En este caso, se está extrayendo una conclusión a partir de una muestra sesgada.

En segundo lugar, el sesgo de una muestra puede surgir de las preguntas que se plantean para obtener los datos. Preguntas tales como: '¿Va con regularidad a la iglesia?' proporcionarán información poco confiable. Puede haber una tendencia en la gente a exagerar su asistencia, ya que, por lo general, se considera como una acción digna. Las palabras 'con regularidad' también causan problemas. El ir dos veces al año, en Navidad y en Semana Santa, es ir con regularidad. También lo es ir dos veces cada domingo. Sería difícil extraer alguna conclusión significativa, si partimos de una pregunta planteada de esa forma. La pregunta debería ser más explícita en la definición de regularidad.

En tercer lugar, el entrevistador puede sesgar la información de la muestra. Por ejemplo, una encuesta realizada en un supermercado acerca de los hábitos de compra puede ser efectuada por un hombre joven, quien entrevista a 50 compradores. No sería de extrañar si la muestra resultante incluyera una elevada proporción de mujeres jóvenes y atractivas.

Más adelante en el curso, se describirán las técnicas de recopilación de muestras o muestreo, que pueden remediar la mayoría de estos problemas.

1.6.4 Omisiones

Las estadísticas que se ocultan pueden ser tan importantes como las que se publican. Un anunciante de televisión afirma que nueve de cada diez perros prefieren la comida para perros Bonzo. El telespectador puede extraer la conclusión de que el 90% de los perros prefieren Bonzo a cualquier otro tipo de comida para perros. Su conclusión podría ser diferente si supiera que:

(a) El tamaño de la muestra era exactamente de diez perros.
(b) Los perros tenían para elegir entre Bonzo y la comida para perros más barata del mercado.
(c) La citada fue la decimosegunda muestra utilizada y la primera en la que nueve perros prefirieron Bonzo.

1.6.5 Errores de Lógica

Las estadísticas permiten extraer conclusiones acerca de las cifras. Usualmente, sin embargo, son las entidades reales las cifras las que son relevantes. Dos de las maneras más comunes de cometer errores de lógica son las siguientes:

En primer lugar, puede que las cifras no correspondan directamente a lo que queremos medir. Por ejemplo, en ocasiones se mide el descontento de los empleados mediante el índice de rotación de personal. Lo que se estudia es el primer aspecto, pero las cifras miden el segundo. Puede que a veces no haya correspondencia entre los dos. Los analistas financieros estudian las cifras de utilidad de las empresas para valorar su rentabilidad. Las cifras de utilidad son, sin embargo, sólo mediciones contables y no pasan de ser unas (ojalá que buenas) aproximaciones de la 'rentabilidad real' de la empresa, la cual es algo complicado tanto de definir como de medir.

En segundo lugar, las conclusiones obtenidas a partir de las cifras no implican necesariamente relaciones de causalidad entre las variables. Por ejemplo, se ha sabido por mucho tiempo que existe una relación entre el salario promedio de los clérigos y el precio del ron. Las dos variables se mueven juntas, lo que puede verificarse estadísticamente. Sin embargo, esto no quiere decir que los clérigos sostengan el precio del ron o viceversa. La explicación es que las variables están vinculadas por un tercer factor, la inflación. Las variables han aumentado juntas a medida que el costo de vida ha aumentado, pero es muy poco probable que exista una vinculación causa-efecto entre ellas. Esta consideración es importante cuando las decisiones se basan en asociaciones estadísticas. Para llevar el ejemplo más allá, el congelar el salario de los clérigos para mantener el precio del ron sería útil si la relación fuera causal, pero no si se trata de una mera asociación.

1.6.6 Errores Técnicos

Ocurren errores cuando se carece de un conocimiento de los aspectos técnicos, aún de los más básicos. Un caso simplista y que se suele citar con frecuencia es el del líder sindical que expone su preocupación por los peor pagados y afirma que no descansará hasta que todos los afiliados cobren un salario superior al salario promedio del sindicato. (Podría ser que en realidad estuviese haciendo una declaración muy sutil).

Otro error sencillo está en el uso inadecuado de los porcentajes. Sería errado suponer que, por ejemplo, un incremento del 20% en la productividad de este año compensa la disminución del 20% del año pasado. Si el índice de productividad dos años atrás era de 100, entonces una disminución del 20% sitúa el índice en 80. El incremento del 20%, entonces, sitúa el índice en 96, es decir, que el índice no ha regresado a su nivel inicial.

1.7 Cómo Detectar Errores Estadísticos

Muchos tipos de errores estadísticos sólo se pueden tratar en el contexto de una técnica cuantitativa determinada, pero existen varias preguntas generales que pueden ayudar a descubrir errores y tretas posibles en la estadística. Estas preguntas deben plantearse siempre que se presenten evidencias estadísticas.

1.7.1 ¿Quién Proporciona la Evidencia?

El derecho aporta una buena analogía. En un caso legal, la opinión de un testigo representa una consideración importante a la hora de evaluar la evidencia. También es de esperar que el abogado defensor no proporcione información que pueda perjudicar a su cliente. Del mismo modo, en la estadística es importante saber quién proporciona la evidencia. Si la persona que proporciona la evidencia tiene interés en que se acepte su conclusión, es necesario prestar especial atención.

No es de esperar que los fabricantes de la comida para perros Bonzo decidan un día declarar lo siguiente: 'Durante mucho tiempo tuvimos la firme creencia de que Bonzo era la mejor comida para perros del mercado. Sin embargo, pruebas recientes realizadas con una muestra aleatoria de 2 000 perros indican que el producto fabricado por la Woof Corporation…'. Por otra parte, es mucho más factible que un informe elaborado por una organización independiente de consumidores aporte evidencias fiables.

1.7.2 ¿Cuál es la Procedencia de los Datos?

Una encuesta de 1992 sobre los hábitos de la población del Reino Unido llevada a cabo por un departamento gubernamental, informa que 'en promedio, la población británica se baña 2.38 veces por semana, en comparación con las 1.15 veces de hace veinte años'. A primera vista, parece ser una evidencia clara, pero, ¿cuál es su grado de fiabilidad?

¿Cuál es la procedencia de los datos? No sería lógico pensar que proviene de observaciones directas. Lo más probable es que se le haya preguntado a la gente. Dado que confesar que uno no se baña con frecuencia podría ser vergonzoso, lo más probable es que las respuestas estén sesgadas. Parece probable que la cifra de 2.38 sea superior a la real. Aún así, todavía es posible comparar los hábitos de la actualidad con los de hace veinte años, pero sólo si se supone que el sesgo en la información sea el mismo ahora que en ese entonces. Pudiera no ser así. ¿Cuál es la procedencia de los datos de hace veinte años? Lo más probable es que provengan de una encuesta cuya estructura fue diferente, sobre una muestra de tamaño diferente, con diferentes preguntas y un entorno social diferente. Por tanto, la comparación con los datos obtenidos hace veinte años también está abierta a dudas.

En este caso, también la exactitud de los datos puede conducir a conclusiones erróneas. La cifra de 2.38 sugiere un elevado nivel de precisión, pero sin respaldo firme debido al método empleado en la recopilación de datos. Cuando las cifras se muestran con varios decimales, conviene dudar de la relevancia de su supuesto grado de precisión.

1.7.3 ¿Supera la Prueba del Sentido Común?

En ocasiones, los especialistas de cualquier campo pueden estar tan absortos en su trabajo que sólo ven los aspectos técnicos y no las cuestiones clave. Las personas ajenas a dicho campo, inhibidas por su carencia de conocimientos técnicos, podrían abstenerse de hacer preguntas de sentido común, con el consecuente detrimento de un proyecto o una investigación. Todo aquello que luzca como opuesto al sentido común debe ser cuestionado.

Un investigador académico estudió la relación entre los ingresos de por vida y la edad en que ocurrió la muerte y descubrió que las dos variables estaban íntimamente relacionadas. Concluyó que la pobreza provoca muertes prematuras.

En primer lugar, alguien podría cuestionar que se está sustentando una relación causa-efecto basada en una asociación estadística. Lo que es más importante, a una persona ajena a la materia quizás le pudiera ocurrir que una vida más larga permite más tiempo para amasar una mayor fortuna y que, por lo tanto, tiene igual (o mayor) validez concluir que la causalidad actúa en el sentido opuesto, es decir, que la muerte prematura de una persona hace que sus ingresos totales de por vida sean menores. El investigador estaba tan sumergido en su trabajo, y puede que hasta prejuiciado por una creencia previa de que la pobreza provoca la muerte prematura, que no fue capaz de aplicar la prueba del sentido común.

1.7.4 ¿Se Ha Cometido Alguno de los Seis Errores Habituales?

En el apartado anterior se describieron los seis tipos de errores estadísticos más comunes. ¿Es posible que se haya cometido alguno de ellos? Revise las seis categorías para comprobar si alguno de ellos es relevante:

(a) **¿Es ambigua la definición?** Podría haberse utilizado un término estadístico (en especial el promedio) susceptible de sugerir más de una interpretación.

(b) **¿Son engañosas las ilustraciones gráficas?** Dé un segundo vistazo, para ver si era posible extraer otras conclusiones. Preste especial atención a la inclusión o exclusión de escalas.

(c) **¿Se ha producido un sesgo en la muestra?** Si se comparan dos muestras, ¿se está comparando lo mismo?

(d) **¿Qué falta?** ¿Se debería haber incluido algún tipo de información adicional, la cual podría haber modificado la conclusión?

(e) **¿Existe un error de lógica?** Puede que las cifras no representen en toda su extensión lo que se pretende medir. Una fuerte vinculación asociativa puede que no sea causal.

(f) **¿Se ha cometido un error técnico?** ¿Se han aplicado en forma correcta las definiciones, las técnicas y los métodos estadísticos? Para poder responder a esta pregunta usualmente se requieren de mayores conocimientos teóricos en cuanto a la materia.

1.8 Observaciones Finales

Esta introducción ha tenido dos propósitos. El primero ha sido presentar algunos conceptos estadísticos, como base para un estudio más detallado de la materia. Todos los conceptos se tratarán con mayor profundidad más adelante. El segundo propósito ha sido el de fomentar un saludable escepticismo y una atmósfera de crítica constructiva, los cuales son necesarios a la hora de ponderar las evidencias estadísticas.

El saludable escepticismo es relevante al enfrentarse, no sólo a los conceptos introducidos hasta ahora, sino a cualquier otro aspecto de la estadística. La probabilidad y las distribuciones son susceptibles a usos incorrectos.

A menudo se comenten errores de lógica al usar la **probabilidad**. Un caso puede ser el envío de un cuestionario sobre métodos de marketing a un grupo de empresas. De las cifras recopiladas de 200 respuestas recibidas, se desprende que 48 de las empresas no están vinculadas al área del marketing. También se determina que 30 de los encuestados son empleados de bajo nivel. ¿Cuál es la probabilidad de que una persona no vinculada al marketing y que además no pertenece a los niveles superiores haya sido quien completó el cuestionario enviado a su empresa? Resulta tentadora la idea de pensar lo siguiente:

$$\text{Probabilidad} = \frac{48 + 30}{200} = 39\%$$

Esto es incorrecto con una certeza casi total debido al doble conteo. Es probable que algunos de los 48 empleados no vinculados al marketing también pertenezcan a los niveles bajos. Si 10 de los que respondieron no estaban vinculados al marketing *y* estaban en un nivel bajo, entonces se concluye lo siguiente:

$$\text{Probabilidad} = \frac{48 + 30 - 10}{200} = 34\%$$

El primer cálculo habría sido correcto sólo en el caso muy poco probable de que ninguno de los de nivel bajo estuviese también fuera del sector del marketing.

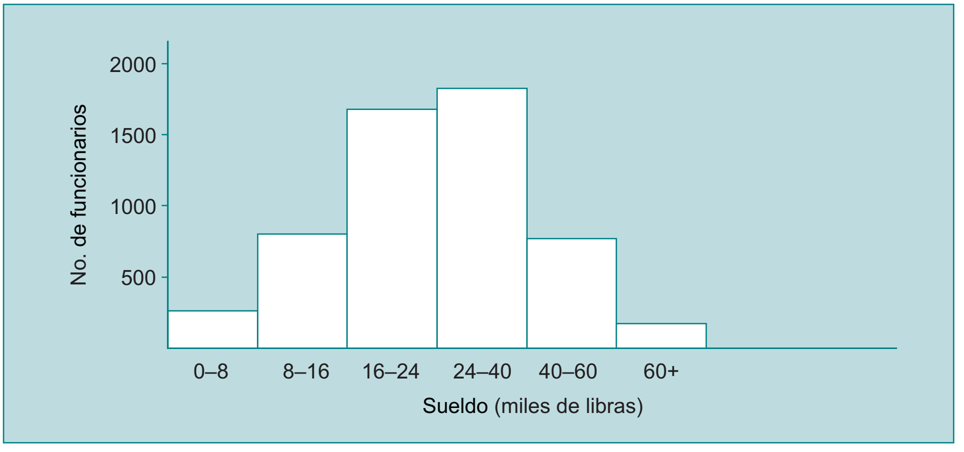

Figura 1.13 Salarios de los funcionarios públicos

Con frecuencia se pueden observar errores gráficos con las **distribuciones**. La Figura 1.13 muestra una distribución observada relacionada con los salarios de los funcionarios públicos en una institución gubernamental. Las cifras dan una impresión errada de la distribución de salarios, debido a que los intervalos entre clases no son todos iguales. Alguien podría pensar que los salarios son superiores a lo que realmente son. Las bandas inferiores tienen una amplitud de £8 000 (0–8, 8–16, 16–24). Las superiores, por su parte, tienen una amplitud mucho mayor. La distribución se debería dibujar con todos los intervalos de la misma magnitud, tal como se muestra en la Figura 1.14.

Los conceptos estadísticos están sujetos al uso incorrecto y a la mala interpretación, del mismo modo que los informes verbales. Se debe estar alerta tanto en el primer caso como en el segundo.

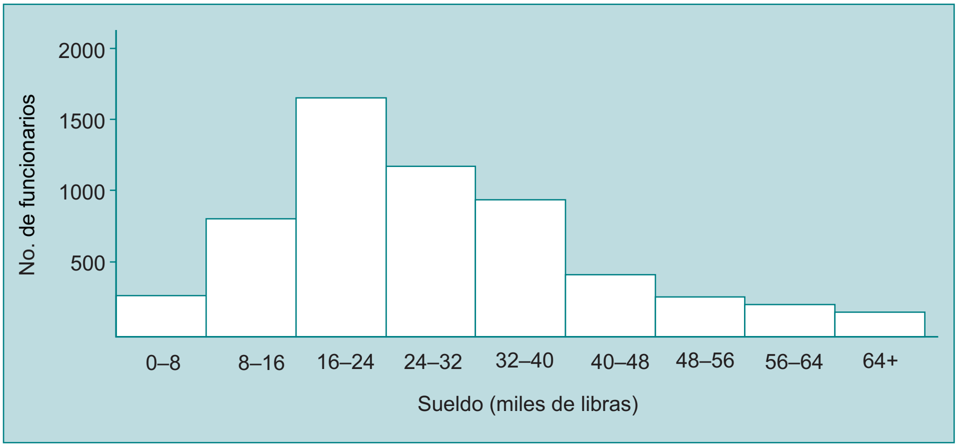

Figura 1.14 Salarios de los funcionarios públicos (corregido)

Preguntas de Repaso

1.1 Una de las razones por las que la probabilidad es importante en la estadística es que si los datos provienen de una muestra, cualquier conclusión que se extraiga no puede ser certera al 100%. ¿Verdadero o falso?

1.2 Se sacó al azar una carta de una baraja y resultó ser un as. No se regresó a la baraja. ¿Cuál es la probabilidad de que al extraer una segunda carta también sea un as?

(a) $\frac{1}{4}$

(b) $\frac{1}{13}$

(c) $\frac{3}{52}$

(d) $\frac{1}{17}$

(e) $\frac{1}{3}$

1.3 ¿Cuáles de las siguientes afirmaciones son verdaderas?

(a) La probabilidad de un suceso es un número entre 0 y 1.

(b) Considerando que nada es completamente certero, ningún suceso tiene una probabilidad de 1.

(c) Los profesionales de la estadística clásica adoptan el punto de vista de que la probabilidad subjetiva no tiene validez.

(d) Los estadísticos bayesianos opinan que sólo la probabilidad subjetiva es válida.

1.4 Si se lanza una moneda normal, existe la misma probabilidad de que caiga 'cara' arriba que de que caiga 'cruz'. Se acaba de lanzar ocho veces y en todas ellas ha salido 'cara'. ¿Cuál es la probabilidad de que en la novena vez el resultado sea 'cruz'?

(a) Inferior a $\frac{1}{2}$

(b) $\frac{1}{2}$

(c) Superior a $\frac{1}{2}$

(d) 1

Las preguntas 1.5–1.7 se basan en la siguiente información:

Las ventas diarias de billetes (en miles de libras) de una estación de tren durante el último trimestre (= 13 semanas = 78 días) se han recogido en un histograma, como se muestra en la Figura 1.15.

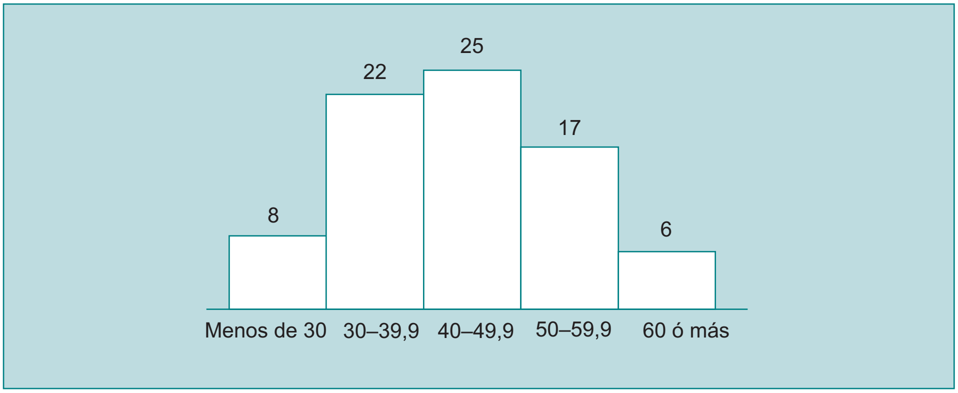

Figura 1.15 **Ventas de billetes de tren**

1.5 ¿En cuántos días fueron las ventas iguales o superiores a £50 000?

(a) 17

(b) 55

(c) 23

(d) 48

1.6 ¿Cuál es la probabilidad de que en un día cualquiera las ventas sean de £60 000 o superiores?

(a) $\frac{1}{13}$

(b) $\frac{23}{78}$

(c) $\frac{72}{78}$

(d) 0

1.7 ¿Cuál es el nivel de ventas que se superó en el 90% de los días?

(a) £20 000

(b) £30 000

(c) £40 000

(d) £50 000

(e) £60 000

1.8 ¿Cuál de las siguientes afirmaciones es verdadera, al hablar de una distribución normal?

(a) Una distribución normal es otra forma de denominar a una distribución estándar.

(b) La distribución normal es un ejemplo de distribución estándar.

(c) La distribución normal es una distribución discreta.

(d) La distribución normal puede ser o no ser simétrica, en función de sus parámetros.

1.9 Una distribución normal presenta una media de 60 y una desviación estándar de 10. ¿Qué porcentaje de observaciones se encontrará en la banda 60–70?

(a) 68%

(b) 50%

(c) 95%

(d) 34%

(e) 84%

1.10 Un puesto de control policial registró la velocidad de los vehículos durante una semana. Las velocidades tenían una distribución normal, con una media de 82 km/h y una desviación estándar de 11 km/h. ¿Qué velocidad fue superada por el 97.5% de los vehículos?

(a) 49

(b) 60

(c) 71

(d) 104

Estudio de Caso Práctico 1.1: Venta de Billetes de Avión

Como primer paso para planificar nuevas instalaciones en una oficina de venta de billetes en el centro de la ciudad, una compañía aérea ha recopilado datos sobre el tiempo que gastan los clientes en las taquillas (el tiempo de servicio). Se utilizó para la investigación a un centenar de clientes y se midió en minutos el tiempo que pasaron en la taquilla de venta. A continuación se muestran los datos.

0.9	3.5	0.8	1.0	1.3	2.3	1.0	2.4	0.7	1.0
2.3	0.2	1.6	1.7	5.2	1.1	3.9	5.4	8.2	1.5
1.1	2.8	1.6	3.9	3.8	6.1	0.3	1.1	2.4	2.6
4.0	4.3	2.7	0.2	0.3	3.1	2.7	4.1	1.4	1.1
3.4	0.9	2.2	4.2	21.7	3.1	1.0	3.3	3.3	5.5
0.9	4.5	3.5	1.2	0.7	4.6	4.8	2.6	0.5	3.6
6.3	1.6	5.0	2.1	5.8	7.4	1.7	3.8	4.1	6.9
3.5	2.1	0.8	7.8	1.9	3.2	1.3	1.4	3.7	0.6
1.0	7.5	1.2	2.0	2.0	11.0	2.9	6.5	2.0	8.6
1.5	1.2	2.9	2.9	2.0	4.6	6.6	0.7	5.8	2.0

Clasifique los datos en intervalos de un minuto de ancho. Elabore un histograma de frecuencias. ¿Cuál es el tiempo de servicio que probablemente será excedido por sólo el 10% de los clientes?

Estudio de Caso Práctico 1.2: JP Carruthers Co.

JP Carruthers Co. (JPC) es una empresa manufacturera de mediano tamaño. Sus cifras de ventas se sitúan en torno a los £220 millones y durante los últimos diez años su nómina ha sido de unos 1 100 empleados. La mayoría de sus ventas se sitúan en el sector automovilístico. La utilidad de JPC del ejercicio anterior fue de £14 480 000. Siempre ha disfrutado de una reputación de confiabilidad, y por lo general, se le considera como una empresa bien administrada.

Con pocas excepciones, la mano de obra directa de JPC, que asciende a cerca de 600 trabajadores, está representada por TWU, el sindicato de trabajadores del transporte. Es práctica común en este sector negociar los beneficios extrasalariales de los empleados de forma global para toda la empresa, pero los salarios de forma separada, según cada tipo de trabajo realizado en cada planta. Durante años, no obstante, esta anticuada práctica ha sido poco más que un ritual. En teoría, ese sistema les ofrece a los trabajadores la oportunidad de expresar sus puntos de vista, pero la realidad es que la fijación de los salarios del primer grupo establece invariablemente el patrón que se seguirá para el resto de los grupos dentro de cada empresa en particular. La línea de ensamblaje de paneles de puerta de JPC fue el grupo clave en la negociación del año pasado. Al ser los primeros de la fila, los salarios que se fijasen para la sección de ensamblaje de paneles de puerta establecerían el patrón de JPC para ese año.

Annie Smith es la capataz de la sección de ensamblaje de paneles de puerta. Hay muchos tipos de paneles de puerta; la principal función de Annie consiste en controlar que cada tipo se fabrique en las cantidades correctas. El trabajo necesario para fabricar los paneles de puerta es prácticamente el mismo, sin importar el tipo de panel en particular. En otras palabras, se trata de un trabajo altamente estandarizado y la remuneración estándar es de 72p por unidad, sin importar el tipo de panel. El trabajo en sí, a pesar de que su naturaleza principal es de ensamblaje, es bastante intrincado y requiere un cierto grado de habilidad.

Las negociaciones del año pasado comenzaron con la habitual protesta del sindicato respecto a las tarifas en general. No obstante, hubo entonces un movimiento inesperado. Esta fue la demanda del sindicato para la línea de ensamblaje de paneles de puerta, según las actas de la reunión:

> 'Vayamos directamente al grano. 72p por unidad es algo diabólico... Una tarifa justa es 80p'.
>
> 'Cada mujer fabrica un promedio de 71 unidades/día. Por lo tanto, los 8p adicionales que queremos equivalen a un aumento promedio de £5.68 por mujer por día...'
>
> 'Es el aumento más pequeño que hemos solicitado recientemente y no aceptaremos menos de 80p.'

(La práctica tradicional de la planta había sido la de calcular la producción en función de la producción promedio diaria. De hecho, a pesar de que la producción de cada persona se contabiliza diariamente, la prima se paga por el rendimiento diario promedio en la semana. La idea es que esto les ofrece a los trabajadores una mejor oportunidad de recuperarse, en caso de tener uno o dos días poco productivos).

La estrategia del sindicato en esta reunión fue una sorpresa. En el pasado, la primera demanda solía ser desproporcionada intencionalmente y ninguna de las partes la tomaba demasiado en serio. En esta ocasión su demanda era similar al tipo de oferta que la dirección de JPC estaba considerando.

En la siguiente reunión de administración luego de la sesión con el sindicato, la dirección de JPC escuchó los siguientes puntos por boca del contador:

(a) La cifra que presenta el sindicato de 71 unidades por persona por día es correcta. Lo he contrastado con el último Informe de Producción. El desglose es así:

La producción promedio semanal en lo que va del año es de 7 100 unidades, por lo que la producción promedio diaria es de:

$$\frac{7\,100}{5} = 1\,420 \text{ unidades/día}$$

El número de mujeres empleadas directamente en esta sección es de 20, por lo que el rendimiento promedio diario es de:

$$\frac{1\,420}{20} = 71 \text{ unidades/día/mujer}$$

(b) La demanda del sindicato equivale a un aumento del 11.1%:

$$\frac{80 - 72}{72} \times 100 = 11.1$$

(c) La mano de obra directa a las tarifas actuales se estima en £26 millones. Si consideramos un aumento global del 11.1%, que por supuesto es lo que hemos de prever, la mano de obra directa anual total aumentaría en cerca de £2.9 millones:

$$£26\,000\,000 \times 11.1\% = £2\,886\,000$$

Con anterioridad a las negociaciones, la dirección había pensado que un 7% podía ser una oferta razonable, ya que estaba aproximadamente al nivel al que había aumentado la productividad y la inflación en los últimos años. En privado, habían fijado en el 10% el límite superior de su oferta final. A este nivel, creyeron que sería conveniente introducir algún tipo de incentivo para mejorar la productividad, aunque no habían desarrollado a fondo los detalles de dicho plan.

Sin embargo, como resultado de la estrategia del sindicato, el equipo de negociación de JPC decidió no vacilar ni un minuto más. Trabajaron hasta tarde y elaboraron su 'mejor' paquete bajo el criterio del diez por ciento. Los puntos principales del plan eran los siguientes:

(a) Mantener los 72p por unidad como tarifa estándar, pero proporcionar una prima de 50p por cada unidad que se produjera por encima de un promedio diario fijado en 61 unidades/persona.
(b) Como el rendimiento promedio por persona por día es de 71, eso significa que en promedio se pagarían 10 unidades de prima por persona por día.
(c) El costo semanal proyectado sería entonces de £5 612:

$$71 \times 0.72 + 10 \times 0.50 = 56.12$$

$$56.12 \times 5 \times 20 = £5\,612$$

(d) El costo actual semanal, entonces, es de £5 112:

$$71 \times 0.72 \times 5 \times 20 = £5\,112$$

(e) Lo que equivale a un incremento promedio de £500 a la semana, ligeramente por debajo del 10%, el límite superior fijado para las negociaciones:

$$\frac{500}{5112} \times 100 = 9.78\%$$

(f) El plan ofrece la ventaja adicional de que el trabajador promedio obtiene 10 unidades de prima de forma inmediata, lo que hace que el plan resulte atractivo.
(g) Debido a que la producción no varía mucho de una semana a otra y a que la mejora más importante debería provenir de aquellos que se encuentran actualmente por debajo del promedio, la mayor parte de los aumentos debería provenir de las unidades a la tarifa más baja de 72p cada una. Probablemente, aquellos que se encuentran por encima del promedio no podrán mejorar mucho. En la medida en que esto ocurra, por supuesto, existe la tendencia de reducir el costo promedio por debajo de los 79p por unidad, que sería el resultado si no se produjese ningún cambio en absoluto:

$$\frac{5612}{71 \times 5 \times 20} = 79.0 \text{ peniques}$$

Llegados a este punto, la administración tenía que decidir entre mostrar todas sus cartas de una vez, o ajustarse al plan original de una oferta inicial del siete por ciento. Se debían tener en cuenta dos aspectos adicionales:

(a) ¿Qué tan buenas eran las tarifas?
(b) ¿Se podría prever realmente un aumento en la productividad, tal y como sugería el plan de la oferta del 9.8%?

Llamaron a Annie Smith, la capataz, para que se incorporase a la reunión, quien proporcionó los siguientes datos:

(a) Algunos, pero no muchos de los trabajadores podían mejorar ligeramente su promedio personal, pero las tarifas eran demasiado ajustadas para que se produjese algún movimiento significativo en la producción diaria.
(b) Esto no quería decir que todos trabajasen al mismo nivel, sino que individualmente todos se encontraban cerca del máximo de su capacidad personal.

(c) Algunos de ellos sí tenían un promedio inferior a las 61 unidades al día. De las pocas que podían mostrar una mejora sostenida, la mayoría se podrían situar en esta categoría de menos de 61 unidades.

Esto lo decidió. JPC decidió ir a la reunión con el sindicato con su 'mejor' oferta del 9.8%. Al día siguiente se hizo la oferta. El sindicato pidió cierto tiempo para considerar la oferta y se fijó la próxima reunión para la tarde del día siguiente.

La mañana del siguiente día, Annie Smith informó de que se había perdido su informe de rendimiento de la producción (*véase* la Tabla 1.4). No sabía quien lo había tomado, pero estaba casi segura de que había sido el representante del sindicato.

Tabla 1.4 Informe de rendimiento de la producción

Semana fiscal: 10	*Centro de gastos: 172*	*Capataz: Smith*
Número del empleado	*Rendimiento diario promedio de esta semana*	*Rendimiento diario promedio acumulado anual*
11	98	98
13	88	89
17	72	76
23	44	43
24	52	50
26	79	78
30	77	79
32	52	52
34	96	96
35	86	87
40	67	69
42	64	66
43	95	98
45	86	88
47	50	53
48	42	41
52	43	44
54	45	46
55	94	97
59	68	70
		Promedio 71
PROMEDIO DIARIO ESTA SEMANA: 1 398		
PROMEDIO DIARIO HASTA LA FECHA: 1 420		

La siguiente reunión con el sindicato no duró más de unos minutos. Un directivo del sindicato expuso su comprensión de la oferta y, luego de serle ratificado que había expuesto los detalles de manera correcta, anunció que el sindicato estaba dispuesto a aprobar el plan y tenía la intención de recomendar su aceptación a todos sus afiliados. También añadió que esperaba que esto sirviera de base para los arreglos con las demás unidades, tal y como venía ocurriendo normalmente y que el conjunto de las negociaciones salariales se podría realizar probablemente en un tiempo récord.

Y hasta allí llegó la cosa. ¿O no? En las mentes de los miembros del equipo de negociación de JPC aún quedaban algunas dudas. ¿Por qué se había dado tanta prisa el sindicato en aceptar? ¿Por qué se había robado el informe de rendimiento de la producción? Mientras hacían conjeturas en cuanto a eso, Annie Smith llamó por teléfono para informar de que habían devuelto el informe de rendimiento de la producción.

Con la esperanza de satisfacer su curiosidad, el equipo de negociación solicitó a Annie que llevase el informe a la oficina. ¿Habían cometido algún error?

¿Era la oferta de JPC realmente del 9.8%? ¿Y si no lo era, cuál era la verdadera oferta?

Estudio de Caso Práctico 1.3: Cartas al Periódico

Las dos cartas adjuntas fueron publicadas recientemente en un periódico. En la primera carta, el Dr. X concluye que los dentistas no deberían suministrar anestésicos. En la segunda carta, el Sr. Y concluye, por su parte, que los dentistas son los anestesistas más seguros que existen.

Comente la evidencia y el razonamiento (según el contenido de las cartas) que lleva a ambas conclusiones.

Peligro en la Silla del Dentista

Estimado Señor: Como anestesista calificado, responsable de un gran número de anestesias dentales, leí (17 de junio) con gran consternación y tristeza acerca del fallecimiento de la Srta. A bajo los efectos de un anestésico.

Me produce una gran preocupación que se le permita a los dentistas suministrar anestésicos. Cualquier tonto puede poner una inyección intravenosa, pero se necesita de una habilidad considerable y de experiencia para poder resolver una emergencia relacionada con anestesia.

Cualquiera que, independientemente de su calificación y de su competencia, suministre anestésicos sin ayuda de ningún tipo, está cometiendo un acto de negligencia criminal. Las asociaciones médicas británicas, así como todas las sociedades de defensa médica, estarían de acuerdo con esto.

Hago un llamado a todos a boicotear la administración de anestésicos por dentistas, bajo cualquier circunstancia.

Atentamente,

Dr. X, Colchester, Essex.

Un Récord de Seguridad Dental Incomparable

Estimado Señor: Con seguridad mucha gente comparte los sentimientos expresados por el Dr. X (Cartas, 25 de junio) sobre el trágico fallecimiento de la Srta. A y eso habla bien de él, pero también lo han confundido.

La Srta. A no fue anestesiada. Fue sedada profundamente con una dosis combinada de drogas que le produjeron una profunda depresión respiratoria, de la cual el practicante no la pudo recuperar.

Al solicitar la prohibición de que los dentistas puedan suministrar anestésicos generales, el Dr. X se adentra en un terreno muy resbaladizo. La posesión de un título médico no confiere por sí misma ninguna inmunidad contra la estupidez o la negligencia, ya que si lo hiciera mucha gente seguiría viva hoy en día.

Si el Dr. X consulta los registros de la oficina de censos y encuestas de población, podrá comprobar que, en general, se producen más fallecimientos relacionados con la anestesia dental cuando el anestesista es un médico calificado que cuando se trata de un dentista.

Si excluimos los servicios hospitalarios (donde todos los anestesistas son médicos calificados, pero donde ocurren aproximadamente el 50% de los casos de fallecimientos), los anestesistas médicos calificados suministran el 36% de las anestesias dentales y a ellos se atribuye el 45% de los fallecimientos. Esto no sólo favorece a los dentistas anestesistas, sino que además demuestra que a cualquiera le puede ocurrir un percance, sin importar su habilidad.

No creo que ni siquiera el Dr. X argumentaría que todas las muertes ocurridas en manos de anestesistas médicos calificados fueran resultado de percances y que todos aquellos casos que hayan ocurrido con dentistas se deban a negligencia.

Sin embargo, estas cifras se deberían poner en la perspectiva correcta. En la práctica odontológica general y en los servicios odontológicos comunitarios se suministran cerca de 1.5 millones de anestésicos al año. En los últimos 15 años, ha habido un promedio de 4 muertes al año. Es un récord de seguridad que no puede ser igualado por ninguna otra forma de anestesia general.

Atentamente,

Sr. Y, (Presidente Electo) Sociedad para el Avance de la Anestesia en la Odontología.