

sesenta y setenta. Muchas de las ideas y estrategias pedagógicas de este texto —el énfasis en las distribuciones muestrales, los seis pasos de la inferencia estadística, los cuatro aspectos de una relación— los aprendí de Dan. De hecho, inicialmente él sugirió que escribiera un texto como su coautor. El tiempo pasó y se retiró antes de que el proyecto avanzara; pero ha sido de mucha ayuda desde entonces. Segundo, un agradecimiento especial a P. Neal Ritchey, mi hermano y colega sociólogo de la Universidad de Cincinnati. Cuando encontré desafíos conceptuales, él siempre estaba allí con su entendimiento y respuestas correctas. Mientras este texto pasaba por borradores proyecto, él aportó espacios dentro de su ocupado horario para leer, criticar y editar para mí, también me ayudó a reunir conjuntos de datos para las aplicaciones en computadora. De verdad aprecio las ventajas de tener un hermano mayor en el mismo campo. Mi amor y agradecimiento para Neal, quien me ha guiado en muchos esfuerzos durante años.

Extiendo un cordial agradecimiento a quienes fueron generosos al brindarme su tiempo y ayuda: Levi Ross y Lynn Gerald por compilar en diskette las aplicaciones de computadora; Lucy Lewis por escribir y reunir los apoyos del profesor; Michael Foti por la extensa edición; Takayo Ashford, Christine Lindquist, Nicole Liddon, Chris McDougal, Jeffrey Mullis, Marilyn Raney, David Sommers, Victoria Smith y Marilyn Wright por ayudar con los ejercicios. Sara Chamberlin, *SPSS, Inc.*; Grant Blank, Universidad de Chicago; y Thomas A. Petee, Universidad Auburn, proporcionaron información sobre paquetes de programas de cómputo. Mary Laska ayudó en la preparación de archivos de datos. Jackie Skeen, Tamalyn Peterson, Helen Dees y Shirley Cottman ayudaron con la mecanografía y la integración de los materiales.

Deseo agradecer a los siguientes y a varios revisores anónimos por sus amplias y constructivas sugerencias: A. Troy Adams, Universidad de Michigan del Este; Jay Alpers, Universidad Palomar; Frank D. Beck, Universidad del Estado de Illinois; William Feinberg, Universidad de Cincinnati; Robin Franck, Universidad del Suroeste; y Surender S. Yadava, Universidad del Norte de Iowa. Mi aprecio se extiende también a mis colegas Kevin Fitzpatrick, Sean-Shong Hwang y Mark LaGory por su consejo y críticas y a Jeffrey Clair, David Coombs, Thomas Edmonds, Guenther Lueschen, Bronwyn Lichtenstein, Earnest Porterfield, George Reinhardt, Joe Schumacher, Ken Wilson, Michelle Wilson, Michael Wrigley y Bill Yoels por sus alentadoras palabras. Y gracias a William Cockerham y Tennant McWilliams por su apoyo, estímulo y consejo.

Por último, agradezco mucho la guía y cooperación de las personas maravillosas de McGraw-Hill. Mi editor, Sally Constable y sus diligentes asistentes, Kate Purcell y Amy Smeltzley, fueron bondadosas con su tiempo y energía, e hicieron que la tarea fuera divertida, así como Kathy Shackelford, quien fue una gran ayuda local. Phil Butcher, Jim Labeots, Leslie Kraham, Jill Gordon y otros ayudaron a que este proyecto se realizara.

FERRIS J. RITCHEY

CAPÍTULO

1

LA IMAGINACIÓN ESTADÍSTICA

Introducción	1
Pensamiento proporcional	3
La imaginación estadística	8
Enlace de la imaginación estadística con la imaginación sociológica	10
Normas estadísticas y normas sociales	10
Ideales estadísticos y valores sociales	12
Estadísticas y ciencia: herramientas para el pensamiento proporcional	13
Estadísticas descriptiva e inferencial	14
¿Qué es la ciencia?	15
Escepticismo científico e imaginación estadística	16
Concepción de los datos	17
El proceso de investigación	20
Insensatez y falacias estadísticas: el problema de los denominadores pequeños	23

Introducción

Un día, cuando el pollo estaba rascando entre las hojas, una bellota cayó del árbol y lo golpeó en la cola. "¡Oh", dijo el pollo, "¡el cielo se está cayendo! Voy a avisarle al rey".

Tomado de *Tomie de Paula's Favorite Nursery Tales*, de Tomie de Paula. Copyright © 1986 por Tomie de Paula. Usado con autorización de G.P. Putnam's Sons, una división de Penguin Putnam Inc.

El pollo hizo algo que todos hacemos de cuando en cuando: poner las cosas fuera de proporción. Mientras ésta es una reacción normal para un libro de cuentos de animales y muchos seres humanos, los peritos estadísticos no deben reaccionar demasiado rápido ni emocionalmente ante dichas situaciones. Un estadístico debe

retroceder y observar desapasionadamente para mantener un sentido claro de equilibrio y proporción.

El campo de la estadística es un conjunto de procedimientos para reunir, medir, clasificar, codificar, computar, analizar y resumir información numérica adquirida sistemáticamente. Un curso en estadística generalmente se percibe como aquel que involucra muchas fórmulas y cálculos. De hecho, algunas operaciones matemáticas están involucradas; pero no constituyen el catalizador de la estadística, y por lo común las computadoras se encargan de esta parte. En realidad, la estadística implica aprender una nueva manera de ver las cosas, adquirir una visión de la realidad basada en el análisis cuidadoso de hechos, en lugar de reacciones emocionales a experiencias aisladas."

El campo de la estadística

Un conjunto de procedimientos para reunir, medir, clasificar, codificar, computar, analizar y resumir información numérica adquirida sistemáticamente.

No todas las búsquedas requieren representaciones exactas y objetivas de la realidad. Los medios populares de comunicación para el entretenimiento —películas, televisión, novelas románticas, entre otros— son, por definición, ficción y fantasía, con personajes y eventos imaginarios. Están diseñados para excitar, alegrar, entristecer o inspirar. De la misma forma, la publicidad llega al mundo entre el hecho y la fantasía, no sólo apelando a la razón sino también a las emociones, para convencer de que con una compra uno se sentirá bien. Las campañas políticas apelan a las emociones de orgullo, patriotismo, temor y odio. Mientras la mayoría de los candidatos son servidores públicos especializados, no todos los políticos se sujetan a los hechos ni se les exige hacerlo; muchos contratan "expertos" para lograr una imagen.

El foro político proporciona un fuerte contraste con la ciencia, pues ésta conlleva un esfuerzo específicamente diseñado para generar un entendimiento más claro de la naturaleza. La ciencia se practica, en la mayor medida posible, de manera independiente de la influencia política o ideológica. El análisis estadístico es una parte vital del método científico. Existe una gran diferencia entre las estadísticas objetivas de las encuestas científicas independientes y las opiniones tendenciosas de encuestadores contratados por políticos ambiciosos. Mientras la meta del personal de las campañas políticas es reforzar la confianza de la gente, empresas independientes intentan estimar la opinión pública. Por ejemplo, para mostrar que un candidato al Congreso va adelante en las preferencias electorales, el personal de la campaña puede contratar una empresa encuestadora que esté dispuesta a hacer preguntas capciosas y preguntar sólo a votantes que han donado dinero. Por supuesto, semejante sondeo revelará un fuerte apoyo, y quizá el personal olvide mencionar a los medios de comunicación que la muestra no era representativa de todos

los electores. Tal manipulación de información numérica hace recordar el dicho de Mark Twain: "Hay mentiras, mentiras detestables y estadísticas."

Si un estadístico profesional dirigiera la misma encuesta, el estudio no oscurecería los hechos ni incluiría preguntas capciosas. En cambio, un estadístico sigue procedimientos cuidadosamente controlados y muestras de la población entera de votantes. Los resultados se presentan con un rango de error y un grado de confianza conocidos, por ejemplo, más o menos 3 puntos porcentuales con 95% de confianza. La estadística trata de lograr una perspectiva equilibrada y una precisión exacta en la recolección y presentación de información.

El principal objetivo de este libro es mostrar que el campo de la estadística versa sobre la obtención de un sentido de proporción exacto con respecto a la realidad; esto significa ver objetivamente las cosas, realizar juicios justos sobre eventos y conductas, dar la cantidad de atención correcta a las cosas que en verdad importan y no distraerse con eventos irrelevantes. Un sentido de proporción ayuda a moderar los sentimientos subjetivos, los sesgos y los prejuicios que distorsionan la percepción de la realidad. Aprender a ubicar las cosas en la perspectiva apropiada requiere imaginación, y en ello descansa el potencial de ver el análisis estadístico como un esfuerzo interesante y disfrutable.

Pensamiento proporcional

El término *proporción* es un concepto matemático relacionado con fracciones y porcentajes. Un buen sentido de proporción sobre un fenómeno requiere más que tener una buena percepción respecto de lo que trata el fenómeno. Entender las proporciones requiere pensamiento proporcional: sopesar la parte contra el todo y calcular la probabilidad de que a la larga ocurra el fenómeno. Tener un sentido de proporción y calcular las proporciones matemáticas son esencialmente lo mismo. Estas últimas son simplemente expresiones precisas de nuestras intuiciones sobre la importancia de ciertos hechos. Calcular una proporción es una manera de medir y evaluar un sentido de probabilidad y significancia respecto de las observaciones realizadas.

Para empezar de manera adecuada, repasaremos brevemente los cálculos básicos de fracciones, proporciones y porcentajes. (Un repaso adicional se proporciona en el apéndice A.) Cada aspecto del trabajo estadístico —desde la medición y la presentación gráfica hasta el cálculo de probabilidades estadísticas— implica trabajar con proporciones matemáticas; por consiguiente, esta revisión ofrece una buena orientación a los cálculos estadísticos.

Las proporciones matemáticas son simplemente *problemas de división que sopesan una parte (el numerador) contra un todo (el denominador)*. Para calcular una proporción, empezamos con una fracción, una manera de expresar qué parte del todo (o número total) constituye una categoría de observaciones.

$$\text{Fracción} = \frac{\text{numerador}}{\text{denominador}} = \frac{\text{parte}}{\text{todo}}$$

Cálculo de una fracción
$\text{Fracción} = \frac{\# \text{ en una categoría}}{\# \text{ en un grupo total}}$
donde # se lee "número de".

Por ejemplo, en un estudio de presos que ocupan la cárcel del condado de Washington, se determina que entre la población total de la cárcel de 149 presos, 112 fueron acusados de delitos relacionados con las drogas (DRD) como la posesión o venta de una sustancia ilegal. ¿Es ésta una gran parte de la población de la cárcel? Si es así, ¿qué dice esto sobre la naturaleza del delito y la aplicación de la ley en el condado de Washington? Para tener un buen sentido de proporción, las dos cifras, 112 y 149, deben integrarse en una fracción:

$$\text{Fracción de presos en la cárcel del condado de Washington acusados de DRD} = \frac{\# \text{ acusados de DRD}}{\text{total de la población de presos}} = \frac{112}{149}$$

Una interpretación más fácil de esta fracción se lograría transformándola en una proporción. Proporción significa *parte de un todo, o parte de la cantidad total o número de observaciones, expresada en forma decimal*. Las fracciones se reducen a proporciones (o expresiones decimales) dividiendo el numerador de una fracción entre su denominador para obtener un cociente. (Un cociente es la respuesta a un problema de división.) Así:

$$p \text{ [de presos en la cárcel del condado de Washington} = \frac{\# \text{ acusados de DRD}}{\text{población total de presos}} = \frac{112}{149} = .7517 \text{ acusados de DRD]}$$

donde *p* simboliza la proporción, la información entre paréntesis describe el total de la población señalada (el denominador) seguido por la característica señalada (el numerador), y el símbolo (#) se lee como "número".

Proporción
Parte de la cantidad total o número de observaciones, expresado en forma decimal.

Cálculo de una proporción
$p \text{ [del grupo total en una categoría]} = \frac{\# \text{ en una categoría}}{\# \text{ en grupo total}} = \text{cociente}$
donde <i>p</i> = proporción de, y el cociente se redondea a cuatro decimales (ejemplo, la diezmilésima más cercana). El cociente siempre tendrá un valor entre cero y 1.

Esta proporción para el condado de Washington es correcta, pero torpemente declarada como "punto siete-cinco-uno-siete" o "siete mil quinientos diecisiete diezmilésimas". Para el público en general, entonces, damos un paso más allá y transformamos esta proporción en la expresión más reconocible: *porcentaje*. Porcentaje significa "por cien", y es igual a una proporción multiplicada por 100. El porcentaje nos dice cuántos de cada 100 presos están acusados de DRD. Así,

$$\% \text{ [de presos en la cárcel del condado de Washington} = p (100) = (.7517) (100) = 75.17 \% \text{ acusados de DRD]}$$

Cálculo de un porcentaje
$\% \text{ [del grupo total en una categoría]} = p (100)$
donde <i>p</i> = proporción del grupo total en una categoría. El cociente siempre tendrá un valor entre 0 y 100 por ciento.

En este momento deberíamos tener la sensación de que el abuso de sustancias prohibidas es un serio problema para la aplicación de la ley en el condado de Washington. De hecho, más de 75 de 100 presos son encarcelados bajo cargos de DRD. Evidentemente, es muy probable y común para una persona encarcelada haber tenido problemas con las drogas. El sistema de justicia en este condado está seriamente agobiado por estos casos.

Las proporciones y porcentajes son medios preferidos para expresar "la parte del todo". Las proporciones siempre tendrán respuestas entre 0 (ninguno) y 1 (todos). De manera similar, los porcentajes siempre oscilan entre el 0 y el 100 por ciento. Aparte de su simplicidad comparada con la forma fraccionaria, las proporciones y porcentajes son útiles para producir rápidamente comunes denominadores para dos o más fracciones. Las proporciones proveen el común denominador 1.00; mientras los porcentajes, el común denominador 100. Por ejemplo, supongamos que comparamos los casos de DRD en las cárceles de los condados de Jefferson y de Washington. Jefferson tiene 42 casos en una población total de 45 presos. ¿Qué

fracción es más grande, 112 de 149 o 42 de 45? Obtenemos un común denominador calculando las proporciones y porcentajes. Para el condado de Jefferson, entonces:

$$p \left[\begin{array}{l} \text{presos en el condado} \\ \text{de Jefferson acusados} \\ \text{de DRD} \end{array} \right] = \frac{\# \text{ de acusados de DRD}}{\text{población total de presos}} = \frac{42}{45} = .9333$$

$$\% \text{ [de presos en el condado de Jefferson = } p(100) = (.9333)(100) = 93.33\% \text{ acusados de DRD]}$$

Los porcentajes permiten ver que, de hecho, la población de la cárcel del condado de Jefferson está más densamente poblada por delincuentes relacionados con drogas que la del condado de Washington (93.33 por ciento contra 75.17 por ciento, respectivamente), aun cuando hay más casos de DRD en el condado de Washington.

Por medio de estos cálculos podemos observar que para cambiar una fracción a una proporción, dividimos el numerador entre el denominador para obtener el cociente "decimalizado". Para cambiar una proporción a un porcentaje, multiplicamos la proporción por 100 moviendo el punto decimal dos lugares a la derecha. Para transformar un porcentaje en una proporción, movemos el punto decimal dos lugares a la izquierda, lo cual simplemente tenemos que dividir entre 100. Para expresar una proporción como una fracción, debemos tener buen dominio sobre los lugares decimales. Si es necesario, repase las posiciones de lugares decimales en el apéndice A. Finalmente, como regla general (con pocas excepciones), redondeamos las proporciones a cuatro lugares decimales a la derecha del punto decimal, y los porcentajes a dos lugares decimales.

Un porcentaje es una manera muy común de estandarizar estadísticas de grupos diferentes. A veces, sin embargo, los porcentajes no transmiten un sentido significativo de proporción. Por ejemplo, ¿cuál es la probabilidad de morir a causa de un relámpago? En 1990, se encontró que la población de Estados Unidos era de 248 709 873 (U.S. Bureau of the Census 1990). Gracias a un meteorólogo se determinó que 74 personas perdieron la vida al ser alcanzados por relámpagos durante ese año.² La proporción y el porcentaje de la población muerta por relámpagos se calcularía como

$$p \left[\begin{array}{l} \text{de la población de} \\ \text{Estados Unidos en 1990} \\ \text{muerta por relámpago} \end{array} \right] = \frac{\# \text{ muertos por relámpagos}}{\text{tamaño de la población total}} = \frac{74}{258\,709\,873} = .00000029$$

$$\% \text{ [de la población de Estados Unidos en 1990 muerta por relámpagos]} = p(100) = .000029\%$$

Así, suponiendo que 1990 es un año típico, la probabilidad de perder la vida por relámpagos es de 29 cienmilésimas de uno por ciento. Esto es difícil de concebir incluso por el individuo matemáticamente astuto. Un denominador de 100 es con-

fuso, cuando menos una persona de cada 100 está en riesgo. La imaginación estadística nos llama poderosamente a encontrar otra manera de interpretar este riesgo.

Otra forma de estandarizar es calculando una tasa, la frecuencia de ocurrencia de un fenómeno en relación con un número "base" especificado de sujetos en una población. El número base se coloca en el denominador para que la tasa pueda representar los casos por mil, por diez mil, por cien mil, por un millón y así sucesivamente. Un número base útil es aquel que claramente especifica "la población en riesgo" para un fenómeno. Con un grupo grande como la población de Estados Unidos, se necesita un número base mayor en lugar del "por 100" usado con los porcentajes. Recuerde que cuando transformamos una proporción en un porcentaje, multiplicamos por 100. De manera similar, podemos multiplicar una proporción por otros múltiplos de 10 para obtener tasas con denominadores mayores.

Cálculo de una tasa

Tasa de ocurrencia = p (un número base útil)

donde p = proporción del grupo total en una categoría y el número base útil es un múltiplo de 10.

Un número base útil para una tasa es aquel que considera la dimensión del conteo de un fenómeno. En este ejemplo contamos a personas fallecidas. Nuestra tasa, entonces, debe presentarse en números enteros con dígitos a la izquierda del punto decimal. Al observar nuestra proporción de .00000029, para obtener un conteo de personas, debemos mover el punto decimal siete lugares a la derecha. El repaso de posiciones de lugares decimales en el apéndice A muestra que esto es equivalente a multiplicar por 10 000 000. Así,

$$\begin{aligned} \text{La tasa de muertes por relámpagos de una población de diez millones} &= \\ (p)(10\,000\,000) &= (.00000029)(10\,000\,000) = \\ 2.9 \text{ fallecimientos por relámpagos de cada } &10\,000\,000 \text{ de personas.} \end{aligned}$$

Este cálculo es explícito y útil. Manifiesta que sólo tres de cada 10 millones de personas mueren por relámpagos cada año. Podemos estimar en 10 millones la población de una ciudad grande (como Nueva York). Al imaginar una ciudad y pensar proporcionalmente, obtenemos la noción de que el riesgo de una muerte por relámpagos es diminuta. Sólo cerca de tres personas en una gran ciudad tienen probabilidad de morir así cada año. (De hecho, si estuviéramos en una ciudad en el desierto, donde rara vez llueve, se reduciría esta cifra.)

Otro cálculo rápido nos permite colocar el numerador de esta tasa a una persona en lugar de 2.9. Esto nos da el número de personas en la población por cada muerte por relámpagos. Una razón de 2.9 a 10 000 000 es igual a una razón de 1 a 3 448 276. Esto se obtiene dividiendo 10 000 000 entre 2.9:

$$\frac{2.9 \text{ muertes por relámpagos}}{10\,000\,000 \text{ de personas}} = \frac{1 \text{ muerte por relámpagos}}{X \text{ personas}}$$

$$X = \frac{10\,000\,000}{2.9} = 3\,448\,276 \text{ personas}$$

donde X es el número de personas en la población por cada persona muerta por relámpago. Así, la probabilidad de morir por relámpagos en un año es, aproximadamente, de 1 en 3½ millones; la población de, por ejemplo, Houston, Texas. De esta forma, la probabilidad de morir por un relámpago es mínima.

Comparación de dos o más grupos de diferente tamaño

Estandarice la fracción usando un común denominador.

Las proporciones tienen un común denominador de 1.

Los porcentajes tienen un común denominador de 100.

Las tasas tienen un común denominador útil seleccionado en múltiplos de 10.

No hemos llegado muy lejos en nuestra discusión introductoria de estadística, y ya hemos identificado la importancia de la comunicación precisa. Las fórmulas matemáticas son bastante estrictas en su forma. Todas las que aquí presentamos tendrán los siguientes elementos:

Planteamiento de respuestas de forma que estimulen el pensamiento proporcional

Símbolo = fórmula = contenidos de la fórmula = respuesta

Observe estos elementos en nuestros cálculos de presos por DRD.

Hacemos una introducción temprana de estos cálculos básicos en el libro, porque tener un sentido de entendimiento sobre la realidad y comprender las matemáticas de proporciones van de la mano. Medidas de "parte del todo" son comúnmente los primeros cálculos realizados en cualquier análisis estadístico. El pensamiento proporcional es una característica básica de la imaginación estadística.

La imaginación estadística

Como ya mencionamos, el objetivo de este texto es proporcionar una nueva visión de la realidad basada en la adquisición de un sentido de proporción y el uso del

pensamiento proporcional. Llamaremos a esta visión la imaginación estadística. El científico social C. Wright Mills (1959)² definió la imaginación sociológica como un conocimiento de la relación del individuo con la sociedad y con la historia. La imaginación sociológica es el reconocimiento de que la conducta individual se rige en función de estructuras sociales más grandes; que la mayoría de las acciones individuales involucra la conformidad con las reglas de la sociedad y no la iniciativa personal; y que, bien o mal, tales reglas se definen dentro de un contexto cultural. La imaginación sociológica implica un detalle aislado (una parte) con respecto a una representación más amplia (el todo): ver el bosque, así como los árboles.

De igual forma, la imaginación estadística consiste en percibir una parte en relación con el todo. La imaginación estadística es una apreciación de cómo un evento usual o inusual, circunstancia o conducta está en relación con un conjunto mayor de eventos similares, y una apreciación de las causas y consecuencias de un evento.

La imaginación estadística

Una apreciación de qué tan usual o inusual es un evento, circunstancia o conducta en relación con un conjunto mayor de eventos similares y una apreciación de las causas y consecuencias del mismo.

Poseer la imaginación estadística es entender que la mayoría de los eventos son predecibles (esto es, ellos tienen una probabilidad de ocurrencia basada en tendencias y circunstancias a largo plazo);³ es tener habilidad para pensar a través de un problema y mantener un sentido de proporción cuando se sopesa la evidencia contra nociones preconcebidas; es poder reconocer eventos muy raros por lo que son, y no por la reacción ante ellos.

Ser estadísticamente *falto* de imaginación es poner las cosas fuera de proporción, para pensar de manera reaccionaria en lugar de proporcional. Por ejemplo, muchas personas se perturbaron con noticias sobre una persona que se había inclinado al caribalismo, como en el caso notorio de Jeffrey Dahmer. Mientras este evento suscitó furia, temor y disgusto, muchos lo vieron como un símbolo del declive moral en Estados Unidos. Semjante noción es reaccionaria. ¡El caribalismo es tan raro ahora como siempre lo ha sido! La imaginación estadística dice: mire esto a la larga, ¿está pasando con frecuencia? ¿Se comprometen muchas personas en esta conducta? ¿Es probable que me convierta en el almuerzo de alguien? De hecho, el incidente de Jeffrey Dahmer fue un caso aislado que involucró a sólo una de entre 250 millones de personas. Ver este evento en su proporción adecuada da la razón a los argumentos sobre la representación más grande de la estabilidad cultural.

Adquirir la imaginación estadística es abrir los ojos a una representación más amplia de la realidad y superar los malentendidos, prejuicios y la estrechez de pensamiento. Por ejemplo, funcionarios de salud pública informan que más de 40 000 personas mueren cada año en accidentes automovilísticos. Confunden el hecho de que los estadounidenses no ven esta importante causa de muerte como un problema de salud pública, relacionado con la seguridad de los caminos y el diseño de auto-

móviles y, por consiguiente, un problema para ser resuelto por políticas gubernamentales; en cambio, el público ve las fatalidades de tráfico como infortunios o fallas individuales. Asumimos que las muertes en el tráfico son el resultado de mala suerte (la víctima se cruzó en el camino de un conductor descuidado), estupidez, imprudencia o descuido (la víctima manejaba con exceso de velocidad o se quedó dormida), mezquindad (demasiado tacaño para comprar nuevos neumáticos) o inmoralidad (la víctima no debió estar bebiendo). ¿Por qué el público no apela a explicaciones individuales pasadas? Una razón es que las muertes y lesiones en el tráfico no golpean frecuentemente a una familia en particular y, por consiguiente, parece que ocurren a "otra persona". Mientras estemos convencidos de que la víctima se lo buscó, nos sentiremos tranquilos de que no nos pasará a nosotros. Por supuesto, nunca beberíamos al mismo tiempo que manejamos, y sólo aceptaríamos donde fuese seguro hacerlo.

La imaginación estadística, sin embargo, nos permite reconocer el efecto a gran escala de este medio de transporte. Miramos la representación amplia de cómo los accidentes de tráfico afectan a la población en contraposición a los individuos. Calculamos las muertes totales y las tasas de mortalidad por millones de millas conducidas usando datos que cubren muchos años. Determinamos qué condiciones inseguras del camino resultan en fatalidades cuando los individuos son descuidados. Por ejemplo, se sabe bien que ocurren más muertes en caminos de dos carriles que en carreteras interestatales de cuatro carriles. De hecho, tomando en cuenta el incremento de automóviles y conductores (traducido en millones de millas manejadas), las tasas de mortalidad en el tráfico han disminuido significativamente desde que el sistema de carreteras interestatal se construyó en los años cincuenta y sesenta. Enfocándonos en el grupo y examinando las circunstancias, además de los individuos, colocamos las muertes en el tráfico en el amplio contexto de salud pública. Sólo entonces empezamos a considerar el valor de la seguridad de otros medios de transporte, como autobuses y trenes subterráneos.

Enlace de la imaginación estadística con la imaginación sociológica

Normas estadísticas y normas sociales

Una visión equilibrada requiere más que cálculo matemático cuidadoso. Por ejemplo, aun cuando se tenga conciencia del número de muertes anuales por accidentes automovilísticos, existen "prejuicios" a favor de este medio privado de transporte. Nos resistimos ante los esfuerzos para su sustitución por los sistemas de tránsito masivo, porque los automóviles encarnan el valor social de libertad individual, fuertemente arraigado entre los estadounidenses. Estamos dispuestos a enfrentar lesión o muerte por libertad y conveniencia.

Cuando los seres humanos usan sus ilustres cerebros para calcular proporciones, porcentajes y otras estadísticas, están simplemente esforzándose por obtener una medida de la realidad. Una estadística, sin embargo, no significa mucho por sí

misma. Un principio importante de la imaginación estadística es que al hacer interpretaciones estadísticas se deben tener en cuenta las circunstancias de un fenómeno, incluso los valores de la sociedad o algún grupo dentro de ella. Los valores sociales pueden llevar a limitar, o quizás incrementar, la respuesta humana a una estadística. En este sentido, cualquier estadística está sujeta a la cultura, es decir, es normativa: su interpretación depende del lugar, tiempo y cultura donde se observa. Una norma social es una idea compartida de la conducta que es apropiada o inapropiada en una situación dada y en una cultura determinada. En una palabra, una norma es una regla; y las normas son peculiares a una sociedad particular, a un periodo de la historia y a la situación específica en que la acción ocurre. Lo que se considera correcto o incorrecto, mucho o poco, depende del lugar y tiempo. Por ejemplo, estar desnudo en la ducha es normal; de hecho, sería peculiar bañarse con la ropa. Estar desnudo en el salón de clases, sin embargo, es un comportamiento desviado (o anormal).

¿Cuándo poco es mucho? Cualquier estadística no tiene sentido si no se establece alguna base de comparación —una norma estadística—. Una norma estadística es una tasa promedio de ocurrencia de un fenómeno. Semejante promedio puede diferir de una sociedad a otra o de un grupo a otro porque cualquier norma estadística es influenciada por normas sociales. Para ilustrar las normas estadísticas y su relación con las normas sociales, comparemos algunas tasas de mortalidad infantil nacional (TMI), el número de niños que mueren en el primer año de vida por cada 1 000 nacidos vivos. La tabla 1-1 presenta la TMI de países seleccionados durante 1992. En Estados Unidos, la TMI fue aproximadamente de nueve muertes por cada 1 000 nacidos vivos.

¿Esta tasa fue alta o baja comparada con la norma estadística? Fue baja respecto de la norma estadística mundial de 1970; pero los estadounidenses no deben sentirse satisfechos por ello. La TMI está estrechamente ligada al desarrollo económico; por consiguiente, la tasa de Estados Unidos se compara más apropiadamente con las normas estadísticas de culturas y economías parecidas, como las de los países industrializados (Japón y los países de Europa Occidental). Resulta que la TMI es bastante alta comparada con la TMI de estas naciones, y los funcionarios de salud pública de Estados Unidos están muy preocupados por ello. Tomado en ese contexto, poco es mucho. Cualquier muerte infantil es significativa para la familia de la víctima; pero esto no inquieta a los funcionarios de salud pública en un país pobre con una alta TMI (así como los estadounidenses no ven las muertes en el tráfico con alarma). Primero, los altos niveles de TMI han perdurado por siglos, haciendo que la norma estadística del país parezca estable. Segundo, las circunstancias culturales —la higiene y el cuidado médico escasos así como la falta de recursos económicos— pueden desafiar enormemente los esfuerzos para reducir esta tasa. Tercero, otras causas de muerte, como el síndrome de inmunodeficiencia adquirida (SIDA) son tan grandes que hacen parecer la TMI bastante baja o simplemente parte de un problema más grande. La forma en que los funcionarios públicos o una sociedad como un todo interpreten una estadística, depende de las circunstancias en un momento dado. Como sugiere este análisis rudimentario de la TMI en la tabla 1-1, la situación cultural influye en la interpretación de los hallazgos estadísticos.

TABLA 1-1 Tasa de mortalidad infantil en naciones seleccionadas en 1992

Nación	Tasa de mortalidad infantil (muertes de niños menores de 1 año de edad por 1 000 nacidos vivos)
Ya industrializada	
Japón	4.4
Islandia	5.5
Suecia	6.2
Alemania	6.7
Canadá	6.8
Gran Bretaña	7.1
Estados Unidos	8.6
Antigua Unión Soviética	20
En proceso de industrialización	
México	38
República Popular China	53
India	91
Haití	105
Etiopía	127
Afganistán	168
El mundo	70

FUENTE: Haub y Yanagishita, 1993.

Para algunas mediciones, como aquellas sobre el desempeño cognoscitivo o conductual, el estado de salud y el logro académico, las normas estadísticas son necesarias incluso para dar sentido a una puntuación. Por ejemplo, con pruebas de coeficiente intelectual (pruebas de CI), las puntuaciones son normadas contra el juicio informado de la comunidad de investigación psicológica sobre lo que constituye la *inteligencia promedio*. Así, las pruebas de CI a menudo son específicamente diseñadas con una norma estadística de 100, número con el que estamos familiarizados y nos sentimos cómodos. Una persona con inteligencia presumiblemente promedio puntúa 100, mientras aquellas que obtienen una puntuación mayor tienen un CI arriba del promedio y quienes presentan una puntuación menor poseen un CI inferior al promedio.

Ideales estadísticos y valores sociales

Una discusión sobre tasas de mortalidad infantil trae a la mente otra distinción que une a la estadística con la realidad social: aquella entre las normas y los ideales. Mientras una norma estadística es un promedio existente, un ideal estadístico es *una tasa de ocurrencia socialmente deseada de un fenómeno, una tasa óptima que se quiere alcanzar*. Los ideales estadísticos reflejan a menudo valores sociales —las ideas compartidas entre los miembros de una sociedad respecto de cómo deberían ser las cosas—. Los valores son las nociones comunes de una sociedad sobre lo que una sociedad

realmente buena tendría en alta estima. En Estados Unidos, por ejemplo, libertad, igualdad, logro, bienestar material, eficacia y nacionalismo son muy valorados (Williams, 1970: 452-500). Estos valores sociales son sólo ideales y nunca se realizan en un sentido puro. Por ejemplo, mientras la libertad individual es altamente valorada, la libertad pura —cada individuo estableciendo sus propias reglas— representa la anarquía. Los valores funcionan como los faros en las playas rocosas. Estas luces sirven como guías, pero alcanzarlas completamente sería arriesgado.

En respuesta a los valores sociales, los ideales estadísticos (tasas de ocurrencia óptimas) a menudo se sustituyen por normas estadísticas. Por ejemplo, la tasa de muerte infantil en Estados Unidos (8.6 muertes por 1 000 nacidos vivos) es un hecho estadístico, y dicha tasa es más alta que la norma para la mayoría de los países ya industrializados, como Japón y Alemania. De esta manera, los funcionarios de salud pública de Estados Unidos quizá señalen las tasas de estos países como un ideal estadístico, una tasa a alcanzar. El público, sin embargo, tal vez no esté dispuesto a aceptar los cambios necesarios para lograr eso, como impuestos más altos y mayor injerencia gubernamental en el cuidado de salud. Los debates sobre ideales estadísticos a menudo revelan conflictos subyacentes y opiniones sobre valores sociales. Tales ideales, entonces, son únicamente eso, y son fuertemente influenciados y restringidos por los valores sociales.

El significado de cualquier estadística a veces depende tan sólo de circunstancias prácticas. Por ejemplo, las normas e ideales estadísticos basados en la biofísica abundan en deportes competitivos, reflejando las limitaciones prácticas de la física. Por ejemplo, ¿cuatro minutos es un tiempo largo? Es un tiempo demasiado largo para completar la vuelta en una carrera automovilística, pero notablemente corto para correr a pie una milla. Un jugador de basquetbol profesional con un porcentaje de tiros libres de 50% está arriesgándose a perder un contrato multimillonario. Sin embargo, un jugador de béisbol de liga profesional necesita dar de hit la pelota aproximadamente 33% de las oportunidades (un promedio de bateo de .333) para ganar el título de bateo en la liga y conseguir un aumento multimillonario. A veces un poco es mucho. La importancia de una estadística depende de las normas estadísticas (promedios), los ideales estadísticos (tasas señaladas como óptimas) y las circunstancias prácticas. La imaginación estadística se emplea para escoger las normas y los ideales estadísticos apropiados con los que se comparan estadísticas y observaciones.

La imaginación estadística, con su conocimiento de la relación entre medidas estadísticas y hechos sociales, requiere un grado de escepticismo —una actitud crítica y suspicaz—. Así como un estadístico se muestra escéptico ante lo que se informa como hecho por aquellos con intereses políticos o económicos, debe aplicarse el escepticismo a la labor de un estadístico, especialmente en el trabajo científico.

Estadísticas y ciencia: herramientas para el pensamiento proporcional

Como hemos mencionado, con la estadística se trata de observar y organizar sistemáticamente información numérica adquirida. La información sistemáticamente adquirida que se organiza siguiendo los procedimientos de la ciencia y la estadística se llama dato o datos.

Las estadísticas y la recolección de datos no son actividades casuales; pero son empresas que requieren un esfuerzo máximo. El análisis estadístico implica precisión; es decir, se refiere a seguir procedimientos, y realizar mediciones precisas y predicciones exactas sobre cómo ocurrirán los eventos en el mundo. Cuando el análisis estadístico se hace de manera apropiada, el analista conoce las limitaciones del razonamiento y de los procedimientos matemáticos, y sabe cuándo las predicciones sobre eventos o conductas son menos que precisas; además, puede expresar el grado de confianza que tiene al hacer una conclusión. En cuánto a esto, el objetivo estadístico es controlar el error. Los errores estadísticos no son equivocaciones.

El error estadístico se refiere a los grados conocidos de imprecisión en los procedimientos utilizados para reunir y procesar información. Controlar el error significa ser tan preciso como sea necesario para reforzar la confianza en las conclusiones derivadas de los hallazgos estadísticos.

Error estadístico

Grados conocidos de imprecisión en los procedimientos utilizados para reunir y procesar información.

El control del error (analizado con detalle en el capítulo 2) representa el segundo tema mayor en este libro y constituye otra faceta de la imaginación estadística. La imaginación estadística no sólo requiere un sentido de proporción acerca de la realidad, sino también la diligencia para mantenerse al tanto de los detalles para minimizar el error.

Estadísticas descriptiva e inferencial

Los datos se reúnen para diferentes propósitos estadísticos. Un propósito de análisis estadístico consiste en tomar muchos datos sobre una categoría de personas u objetos, y resumir esta información en pocas cifras matemáticas exactas, tablas o gráficas. Este primer paso en estadísticas se llama estadística descriptiva.

La estadística descriptiva explica cuántas observaciones fueron registradas y qué tan frecuentemente ocurrió en los datos cada puntuación o categoría de observaciones. Por ejemplo, datos tomados de 291 encuestados muestran que 40 por ciento son varones y tienen una edad promedio de 21 años, siendo el más joven de 19 y el más viejo de 51 años. La estadística descriptiva es empleada tanto por científicos como por encuestadores, analistas de mercadotecnia, proyectistas urbanos y muchos otros profesionales. Estos cálculos informan al público sobre qué productos adquirir, a qué políticos creer, qué acciones comprar, qué automóviles son los más confiables, a qué edad se recomiendan las revisiones físicas anuales, y asuntos por el estilo. La estadística descriptiva también es utilizada por científicos como un primer paso en el análisis de hipótesis de investigación científica, que es la tarea de la estadística inferencial.

Un segundo propósito del análisis estadístico es extraer conclusiones sobre las relaciones matemáticas entre las características de un grupo de personas u objetos. Por ejemplo, podríamos investigar si los estadounidenses que están mejor educados tienden a creer menos que el diablo existe. Este tipo de análisis se denomina estadística inferencial y se calcula para mostrar relaciones de causa-efecto, así como para probar hipótesis y teorías científicas. (Inferir quiere decir sacar conclusiones sobre algo.) La mayor parte de este texto trata sobre estadística inferencial. Entender los principios básicos de la ciencia es imprescindible para comprender la estadística inferencial; por lo tanto, repasaremos estos principios.

¿Qué es la ciencia?

La ciencia es un método sistemático para la explicación de los fenómenos empíricos. Empírico significa observable y medible. Los fenómenos (forma plural derivada de la palabra en latín *phanomenon*) son hechos, situaciones, eventos, circunstancias, o, simplemente, "cosas que existen naturalmente". Los fenómenos empíricos, entonces, son cosas que pueden observarse y medirse, como condiciones naturales, procesos, eventos, situaciones, objetos, grupos de personas, conductas, pensamientos, creencias, conocimientos, opiniones, emociones y sentimientos.

No todo es medible y observable. Por ejemplo, si existe vida después de la vida no se observa fácilmente, aunque cerca del 70 por ciento de adultos estadounidenses afirma creer en ello. Además, muchas cosas intangibles, como las emociones, sentimientos y creencias deben medirse de forma indirecta. En las ciencias sociales tales mediciones indirectas incluyen encuestas que sondean opiniones, conocimiento, actitudes e incluso conducta. Los científicos físicos también usan mediciones indirectas; por ejemplo, los físicos indirectamente observan neutrinos, partículas subatómicas tan diminutas y rápidas que comúnmente atraviesan la tierra sin golpear con algo. (Millones de ellas están atravesándola ahora mismo.) En ocasiones, un neutrino desplaza una molécula de agua, liberando energía observable, y este efecto puede medirse. Un aspecto importante de la expansión de la ciencia está en encontrar nuevas formas de medir con exactitud los fenómenos que no son visibles al ojo humano. Microscopios, máquinas computarizadas de rayos X y sismómetros, así como instrumentos de investigación, son las herramientas que los científicos emplean para extender el alcance de sus capacidades de medición.

El propósito de la investigación científica. El objetivo principal de la ciencia es explicar los fenómenos. Una explicación científica está basada en procedimientos estrictos, y se llama teoría. Una teoría científica es un conjunto de aseveraciones interrelacionadas y lógicamente organizadas, que explican un fenómeno de especial interés, y que han sido corroboradas a través de la observación y el análisis. Las teorías describen situaciones y cómo funcionan éstas. La colección de ideas que constituyen una teoría, se prueba contra los hechos observados. Una teoría se "corroborada" cuando sus ideas predicen con éxito estos hechos observables. Una teoría no es un hecho en sí misma; es una explicación bien organizada de hechos. Cuando un fenómeno se entiende mejor, una teoría se modifica y se afina para aumentar su poder de predicción; así, su desarrollo es un proceso acumulativo que ocurre durante un periodo largo.

Una teoría científica adecuada logra dos cuestiones. Primero, proporciona un sentido de comprensión sobre un fenómeno: cómo, cuándo, por qué y bajo qué condiciones sucede; descrito de manera sencilla, da sentido a las cosas. Segundo, una teoría nos permite realizar predicciones empíricas, respondiendo a la pregunta de bajo qué condiciones y a qué grado un fenómeno ocurrirá. Tales predicciones son posibles porque los cambios en un fenómeno se relacionan con cambios en otros fenómenos. Por ejemplo, predecimos una mayor oportunidad de lluvia cuando la humedad atmosférica aumenta, o un incremento en la tasa de delitos en una comunidad cuando hay crisis económica.

Escepticismo científico e imaginación estadística

La ciencia *requiere* que sus ideas resistan la prueba de predecir observaciones. Los científicos especializados son escépticos; mantienen una actitud crítica y suspicaz; están dispuestos a tolerar incertidumbre, por lo que no son demasiado rápidos para obtener conclusiones. Un escéptico vacila en creer algo simplemente porque sus amigos de confianza, los medios de comunicación masiva o personas en posiciones de autoridad, como líderes gubernamentales o incluso sus padres, lo consideran verdadero. Un vistazo a la cultura popular, sobre todo a las ideas lanzadas en los medios de comunicación masiva, sugiere que la mayoría de las personas son altamente crédulas—inclinadas a creer—incluso en ausencia de evidencia o en presencia de evidencia contradictoria. El renombrado científico Carl Sagan, un portavoz del valor de la ciencia, observó que muchas personas son muy lentas para "suspender el escepticismo". Por ejemplo, él notó cómo las personas crédulas estaban cayendo en el engaño de los "círculos de la cosecha", al creer, durante 15 años, que los enormes y elegantes pictogramas descubiertos en campos de cultivo ingleses habían sido dejados allí por extraterrestres. (Un par de bribones llamados Bower y Chorley finalmente confesaron ser los autores del engaño.) Sagan argumentaba que no somos lo suficiente escépticos ante mucho de lo que se presenta como hecho sólo por alegato. Él exhortó a tomar la ciencia más en serio, porque el proceso científico se diseña especialmente para separar el hecho de la ficción (Sagan 1995a, 1995b). Además, afirmó que una sociedad que fomenta el aprendizaje de la ciencia producirá ciudadanos mejor informados. Sagan señaló lo siguiente:

En la universidad [...] empecé a aprender un poco sobre cómo funciona la ciencia [...] qué tan rigurosas deben ser las normas de evidencia [...] cómo nuestros prejuicios pueden afectar nuestra interpretación de la evidencia, cómo los sistemas de creencias ampliamente [...] apoyados por las jerarquías políticas, religiosas y académicas a menudo resultan no sólo estar ligeramente errados sino grotescamente equivocados...

Los principios del escepticismo no exigen un grado avanzado de dominio [...] Lo que la ciencia pide es que empleemos los mismos niveles de escepticismo que usamos para comprar un automóvil usado...

Sin embargo, las herramientas del escepticismo por lo común no están disponibles para los ciudadanos de nuestra sociedad. Casi nunca se mencionan en las escuelas, incluso en la presentación de la ciencia, su practicante más ardiente [...] Nuestras políticas, economías, publicidad y religiones (de la nueva o de la vieja era) están inmersas en la credulidad. Un escéptico podría sugerir que aquellos que tienen algo que vender,

quienes desean influir en la opinión pública, o aquellos que están en el poder, poseen un velado interés en desmotivar el escepticismo* (Sagan, 1995b:10-13).

Las explicaciones científicas basadas en la observación, los procedimientos estrictos y el escrutinio colectivo de la comunidad científica a menudo contradicen el sentido común, así como las ideas propuestas por líderes políticos. Esto no significa que la ciencia abandone el sentido común. La ciencia utiliza el sentido común informado, *que es evaluado y doblemente verificado contra datos cuidadosamente recogidos*. ¡El sentido común desinformado es demasiado común! El escepticismo científico requiere aprender habilidades procedimentales y desarrollar una actitud de cuestionamiento. De igual manera, poseer la imaginación estadística involucra aprender habilidades (por ejemplo, cómo calcular probabilidades y pensar proporcionalmente) y estar listo para preguntar si un fenómeno observado es razonable.

Al mismo tiempo, sin embargo, la ciencia tiene limitaciones. Primero, se restringe a exámenes de fenómenos empíricos—observables y medibles—; la fe, no la ciencia, debe resolver, por ejemplo, la pregunta de si Dios y la salvación del alma existen. Segundo, muchos sondeos, objetivamente basados en argumentos científicos, carecen de apoyo político o de los contribuyentes. Por ejemplo, la investigación revela que la pobreza en Estados Unidos se reduciría extendiendo los programas gubernamentales de ayuda familiar, como capacitación para el trabajo y servicios de cuidado infantil. Estos "programas de asistencia", sin embargo, son costosos y a menudo carecen de apoyo del contribuyente; la reciente legislación da simplemente un tiempo limitado a los destinatarios de la asistencia pública para resolver estos problemas por sí mismos. Una tercera limitación de la ciencia es que provoca dilemas éticos y resistencia ante su aplicación. Por ejemplo, un economista podría elaborar un argumento convincente de que la eutanasia o "ruerte por piedad" ahorraría miles de millones de dólares en gastos médicos destinados a enfermos terminales. Obviamente, muchos cuestionarían tal argumento, no con base en la cuestión monetaria, sino en la moral. Hay más que explicar acerca de la existencia humana que de costos.

La ciencia no tiene todas las respuestas, y los científicos deben mantenerse escépticos de las respuestas que poseen. Cuando se explica la realidad empírica, sin embargo, el método científico constituye el mejor enfoque. Un rasgo importante del método científico es el análisis estadístico.

Concepción de los datos

Variables y constantes. *Fenómenos medibles que varían (cambian) a través del tiempo o que difieren de un lugar a otro o de un individuo a otro, se denominan variables. Las variables son características de los sujetos (estudiantes, personas sin hogar, habitantes de St. Louis, ratas de laboratorio) u objetos (edificios, árboles, inundaciones, bacterias, delitos) bajo estudio.* (De aquí en adelante, se empleará el término *sujeto*

* Carl Sagan, *The Demon-Haunted World*, Copyright © 1995 por Random House, reimpresso con autorización del editor.

para designar tanto a personas como objetos.) Por ejemplo, al estudiar a los individuos, se notarían diferencias en las variables de edad, peso, altura, rasgos de personalidad, raza y estatus socioeconómico.

Variable

Fenómeno medible que varía (cambia) a través del tiempo, o que difiere de un lugar a otro o de un individuo a otro.

Se utiliza el término variación para referirse a cuánto difieren las mediciones de una variable entre los sujetos en estudio. Se comparan las diferencias en la variación entre grupos. Existe mucha variación, por ejemplo, en las edades de estudiantes universitarios en las grandes ciudades, quizás van desde 17 hasta 70 años. En contraste, esta variación en universidades tradicionales en pequeños "pueblos aledaños a un campus" es comúnmente más pequeña, de 17 a 25 años.

Algunas variables muestran poca o ninguna variación dentro de un grupo, como las edades de los alumnos de primer grado. Las características de los sujetos en estudio que no varían se llaman constantes. A veces, de manera intencional "mantenemos constantes las variables". Por ejemplo, en un experimento sobre los efectos de las bebidas alcohólicas en la conducta del automovilista, usaríamos sujetos de más o menos el mismo peso, porque se sabe que las personas más ligeras se embriagan con mayor rapidez que las más pesadas. De esta manera, la reducción del tiempo de reacción al manejar se atribuiría a la cantidad de alcohol consumida en lugar de a las diferencias en peso. "Manteniendo el peso constante", eliminamos sus efectos en la conducta del automovilista; puesto que el peso no variaba, una variación en el peso no podría explicar los resultados del experimento. Manteniendo constantes el peso y cualquier otra variable que afecte la embriaguez, somos capaces de aislar los efectos de consumo de alcohol en la conducta de quien maneja.

La variable dependiente y las variables independientes que la explican. Por lo común, al recolectar datos, nuestro propósito consiste en investigar una sola variable que es de especial interés para nosotros. Queremos saber qué provoca un incremento o disminución en la cantidad de esta variable. ¿Qué causa dicha "variación"? ¿Cuáles son sus puntuaciones dependientes? Esta variable de principal interés se denomina variable dependiente, la variable cuya variación queremos explicar. Por ejemplo, los años sesenta se caracterizaron por la violencia urbana, con disturbios en más de 40 ciudades durante un periodo de tres años. En un esfuerzo por entender y prevenir los disturbios, la Comisión Nacional de Asesoría sobre Desórdenes Civiles (1968) se formó para dirigir un estudio científico. La incidencia de conducta tumultuaria fue la variable dependiente. La comisión quiso explicar por qué los disturbios ocurrieron en algunas ciudades pero no en otras.

Las variables que se sospechaba estaban relacionadas con un incremento o disminución en la conducta tumultuaria también se midieron. Tales variables incluyeron la tasa de pobreza en las comunidades, el número de quejas por brutalidad

policíaca, las perturbaciones raciales en las semanas previas a un disturbio y el número de "simpatizantes comunistas" conocidos en una ciudad. Estas variables predictoras que están relacionadas o que predicen la variación en la variable dependiente, se conocen como variables independientes. La tabla 1-2 distingue las características de las variables independientes y de las variables dependientes.

La comisión sobre desórdenes civiles examinó ciudades como sus sujetos de investigación. El equipo de investigación de la comisión encontró que la incidencia de los disturbios (la variable dependiente) estaba relacionada con las variables independientes como el porcentaje de familias que viven en la pobreza, la suficiencia de programas de asistencia social, el grado de participación gubernamental de las minorías, la ocurrencia de "incidentes de elevada tensión" y, especialmente persistentes, informes de brutalidad policíaca. Para mostrar que existe relación entre la incidencia de brutalidad policíaca y la incidencia de disturbios, se dice que las ciudades con elevada brutalidad policíaca también tendieron a sufrir muchos disturbios. Esta afirmación y otras similares involucran otras variables independientes que conforman una teoría de protesta de la conducta tumultuaria. Esta teoría propuso el argumento de que las personas provocan disturbios en respuesta a las acciones policíacas opresivas y debido a la frustración por escasos servicios del gobierno. Las ideas y datos estimulados por esta teoría finalmente auspiciaron cambios en políticas gubernamentales locales y una reducción en los desórdenes civiles (Johnson, 1973: 376).

Los hallazgos de la comisión refutaron algunos mitos, creencias ampliamente sostenidas que son falsas y que estaban muy arraigadas. Específicamente, se desacreditó la teoría de la conspiración comunista, un argumento político que sostenía que los disturbios eran parte de una revolución organizada para derrocar al gobierno de Estados Unidos (Johnson, 1973:376). ¿Por qué de manera tan inmediata las personas creyeron que los disturbios representaban una conspiración comunista? Los mitos a menudo surgen de explicaciones de sentido común reforzadas por eventos aislados o esporádicos y por la retórica política que aviva los miedos del electorado. La violencia urbana de los años sesenta ocurrió durante un periodo de cambio social rápido e

TABLA 1-2 Posibles relaciones entre las variables independientes y dependientes

Variable independiente		Variable dependiente
Causa	→	Efecto
Predictor	→	Resultado
Estímulo	→	Respuesta
Intervención (acción tomada)	→	Resultado
Correlación: cambio en una variable	↔	Cambio asociado en otra variable

incierto. En el frente doméstico había un movimiento de los derechos civiles organizado con minorías raciales, especialmente afroamericanos, que exigían la eliminación de la discriminación en la contratación para empleos, en las escuelas y en el uso de instalaciones públicas. En la escena mundial existía al mismo tiempo una "guerra fría" entre los países capitalistas de occidente y los países comunistas; de estos últimos, sobre todo, la antigua Unión Soviética y la República Popular China, cuyos gobiernos realizaron llamados abiertos a las armas y buscaron infiltrar espías en Estados Unidos, los cuales exaltarían a los pobres y a las "minorías reprimidas" para sublevarse. Ocasionalmente, organizaciones socialistas locales distribuyeron folletos en los que se hacía el llamado a un levantamiento armado para eliminar la opresión racial. Tales protestas, aisladas pero emocionalmente cargadas, facilitaron a los políticos jugar con los miedos públicos. Muchos políticos intentaron desacreditar a sus oponentes que apoyaron el cambio racial acusándolos de tener inclinaciones comunistas. En esta atmósfera, la ocurrencia de disturbios en barrios habitados por minorías pobres a lo largo de Estados Unidos parecía, a muchas personas, un resultado verosímil de la conspiración comunista. La tendencia del público para creer esta explicación tan pronto fue reforzada por políticas raciales, así como por la incertidumbre y el temor que acompañaron los vertiginosos cambios sociales.

Cuando así sucedió, los hechos demostraron que era sumamente difícil encontrar a los comunistas entre los participantes del disturbio. Además, no había ninguna diferencia en el número de simpatizantes comunistas en ciudades donde los disturbios ocurrieron y en ciudades donde no los hubo. El argumento de la conspiración comunista fue refutado, no fue apoyado por datos y no resistió el escrutinio del análisis estadístico.

Una teoría científica es un argumento organizado que debe ser corroborado por la evidencia empírica. Una teoría se "corrobor" cuando sus ideas predicen exitosamente mediciones observables.⁵ Cuantos más datos se adquirieran, más se modificarán y refinarán las teorías para mejorar su poder de predicción y su sentido de comprensión.

El proceso de investigación

El proceso de investigación implica organizar ideas en una teoría, realizar predicciones empíricas que apoyen la teoría, y entonces, reunir datos probatorios de tales predicciones. El proceso de investigación es acumulativo, es decir, un proceso continuo de acumulación de conocimiento. El proceso de investigación científica comprende siete pasos que serán enseñados en diferentes etapas: del 1 al 3 son los principales temas en los cursos de teoría en ciencias sociales; los pasos 4 y 5 se cubren en cursos de metodología, y los pasos 6 y 7 se enseñan en cursos de estadística. Los siete pasos son los siguientes:

1. *Especifique la pregunta de investigación.* Planteamos una pregunta e identificamos la variable dependiente. Por ejemplo, podemos preguntar: ¿por qué están ocurriendo disturbios en algunas ciudades?

2. *Revise la literatura científica.* Hacemos esto para asegurarnos de que no se desperdicien tiempo y dinero recolectando datos que ya existen. Buscamos la "fron-

tera del conocimiento", los límites exteriores de lo que ya ha sido aprendido, por ejemplo, sobre los disturbios. La investigación bien informada y publicable extiende el conocimiento más allá de las fronteras.

3. *Proponga una teoría y formule una hipótesis.* La teoría involucra la organización de ideas en una forma lógica que pueda explicar la variación en la variable dependiente. Al desarrollar una teoría, identificamos las variables independientes y hacemos declaraciones predecibles respecto de cómo pensamos que afectan a la variable dependiente, asumiendo que la teoría es comprensible. Tales predicciones se llaman hipótesis. Una hipótesis es una predicción sobre la relación entre dos variables en ella se afirma que los cambios en la medida de una variable independiente corresponderán a cambios en la medida de una variable dependiente. (Como analizaremos en capítulos posteriores, una hipótesis también indica la dirección de una relación; si la variable independiente se relaciona con un incremento (dirección positiva) o una disminución (dirección negativa) en la variable dependiente.)

Las hipótesis se generan o "motivan" por la teoría—ideas probadas que han sido encontradas en la literatura científica—, con modificaciones innovadoras del investigador. La teoría nos lleva a esperar ciertos resultados observados de los datos. Si estos resultados ocurren, la teoría se corrobora. Por ejemplo, la teoría de la protesta en la conducta de disturbio motiva la siguiente hipótesis:

H₁: Las ciudades con alta incidencia de brutalidad policiaca (variable independiente) están sujetas a tener una elevada incidencia de desórdenes civiles (variable dependiente).

En contraste, la teoría de la conspiración comunista en la conducta de disturbio da origen a la siguiente hipótesis:

H₂: Las ciudades con un gran número de comunistas (variable independiente) están sujetas a tener una elevada incidencia de desórdenes civiles (variable dependiente).

Además de probar la hipótesis, teorizar implica la organización lógica de ideas y filosofías. La teoría, basada en la revisión de la literatura, también guía en la selección de variables "control". Por ejemplo, al medir desórdenes civiles, debemos controlar la tasa de delitos. Esto nos asegura que las ciudades con una alta incidencia de disturbios no son simplemente ciudades con mucha criminalidad, en las cuales la tasa de casos de delitos—no sólo la brutalidad policiaca—, explicaría parte de la incidencia en desórdenes civiles. El concepto de control estadístico se estudia en el capítulo 2.

También debemos notar que no todos los estudios científicos emplean teoría. Mucha investigación se lleva a cabo para resolver problemas prácticos inmediatos o para explorar nuevos fenómenos sobre los que se conoce tan poco que formular una teoría sería imposible. Tales estudios se llaman estudios exploratorios. Por ejemplo, alguien que explora cuestiones privadas en Internet empezaría con ideas y preguntas vagamente organizadas.

Hipótesis

Predicción sobre la relación entre dos variables; en ella se afirma que los cambios en la medida de una variable independiente corresponderán a cambios en la medida de una variable dependiente.

4. *Selección un diseño de investigación.* En el diseño de investigación se detalla cómo se medirán, muestrearán y reunirán los datos. Los métodos comunes de la ciencia social incluyen la observación directa del comportamiento, el experimento de laboratorio, la encuesta, el análisis de contenido en los medios de comunicación y el análisis de datos existentes o "secundarios" (como informes policíacos y censos de población).

5. *Recolecte datos.* Ésta es normalmente la parte más costosa de la investigación. Se trata de "entrar en el campo" para informar a las personas sobre el estudio y recolectar datos utilizando el plan desarrollado en el paso 4. También es una de las partes más agradables de la investigación, pues permite al investigador salir de la oficina y conocer nuevas y, a menudo, interesantes personas.

6. *Analice los datos y saque conclusiones.* Es aquí donde entra el análisis estadístico, tema principal de este libro. Las hipótesis se prueban mediante la comparación de observaciones con predicciones teóricas. En el ejemplo de los disturbios, los datos recolectados por la Comisión Sobre Desórdenes Civiles apoyaron la hipótesis 1 y refutaron la 2, otorgando mayor credibilidad a la teoría de protesta.

7. *Difunda los resultados.* Difundir significa diseminar ampliamente y compartir. Los hallazgos científicos se comparten con dos tipos de "audiencias": el público y la comunidad científica.

Las audiencias públicas no sólo incluyen a los ciudadanos sino también a grupos políticos y empresariales, religiosos, de caridad y educativos. Los investigadores pueden exponer en foros públicos a manera de conferencias de prensa, entrevistas, reuniones oficiales en la ciudad, reuniones comunitarias y clases de bachillerato. Tales charlas deben ser sencillas conceptual y estadísticamente.

Para la audiencia científica, la difusión de hallazgos de investigación consiste en presentar los descubrimientos en conferencias científicas y libros que serán publicados o, más comúnmente, en artículos cortos en revistas especializadas. La publicación de la investigación es un arduo proceso de revisión entre pares —un sistema de comprobaciones y acuerdos—, que idealmente aumenta al máximo la probabilidad de que un trabajo publicado sea preciso e imparcial. Un manuscrito científico sigue un procedimiento estricto. Cuando se completa, se somete al juicio del editor de una revista especializada en la materia, quien, a su vez, envía copias sin identificación del autor a otros científicos competentes en la materia. Esta revisión "ciega" minimiza el prejuicio personal; obliga a los revisores, sin embargo, a que sean altamente escépticos con el manuscrito. Ellos escrutan cada detalle buscando fallas lógicas, interpretaciones tendenciosas, muestras sin sondeo, medición pobre o análisis y conclusiones equivocados. Si varios revisores están de acuerdo en que la investigación es sólida y que contribuirá al avance del conocimiento, el

editor podría aceptar la publicación si el espacio de impresión está disponible. Las principales revistas, en la mayoría de los campos, son sumamente selectivas, pues publican sólo una de cada 10 solicitudes. Este proceso asegura que la investigación seleccionada para su publicación alcanza estándares profesionales. Los investigadores que publican regularmente son practicantes de la ciencia altamente calificados.

Los revisores de los manuscritos son, y deben ser, muy críticos; no del investigador, sino de la calidad de los procedimientos científicos que éste siguió. Científicos, investigadores serios de mercadotecnia, encuestadores y otros que asimilan el trabajo estadístico desarrollado cuidadosamente, aprenden a respetar y apreciar el escrutinio de sus compañeros analistas. Carl Sagan expresó su preocupación por la falta de escepticismo que muchas personas muestran, pero también señaló con orgullo los excelentes esfuerzos de aquellos comprometidos en realizar el trabajo de la ciencia de manera correcta. El capítulo 2 se enfoca en el control del error y en cómo esto exige un compromiso concienzudo para hacer el mejor esfuerzo al dirigir el trabajo estadístico.

⊗ INSENSATEZ Y FALACIAS ESTADÍSTICAS ⊗

El problema de los denominadores pequeños

Se debe tener cuidado al interpretar proporciones y porcentajes basados en grupos sumamente pequeños; los números pequeños en la línea base en reportes de cambio de porcentaje son una particular fuente de confusión. La tabla 1-3 presenta un ejemplo ficticio de lo que típicamente ocurrió al principio de la epidemia del SIDA.

El cambio de porcentaje se calcula como sigue:

$$\text{Cambio de porcentaje} = \frac{\# \text{ al tiempo 2} - \# \text{ al tiempo 1}}{\# \text{ al tiempo 1}}$$

TABLA 1-3 Cambio del porcentaje en el número de nuevos casos de SIDA informados en un condado, de 1988 a 1989 por género (datos ficticios)

Género	Número de nuevos casos en 1988	Número de nuevos casos en 1989	Cambio de porcentaje de 1988 a 1989
Hombres	78	104	33%
Mujeres	4	7	75%
Total	82	111	35%

La tabla muestra que el *incremento de porcentaje* en la incidencia del SIDA fue mayor para mujeres que para hombres entre los dos años. Tales estadísticas se reportaron a menudo como evidencia de que la epidemia estaba extendiéndose mucho más rápidamente entre las mujeres que entre los hombres, sugiriendo que el SIDA de repente se había vuelto una enfermedad "femenina". De hecho, en 1989 sólo 7 nuevos casos aparecieron entre mujeres, comparados con 104 en hombres. El aparente fenómeno "femenino" se debió al problema de un denominador pequeño. En semejante situación, un buen estadístico simplemente reportaría que había muy pocos casos de mujeres para realizar comparaciones significativas.

Fórmulas en el capítulo 1

Al mostrar el trabajo cuando se efectúan cálculos:

$$\text{Símbolo} = \text{fórmula} = \text{contenido de la fórmula} = \text{respuesta}$$

Al calcular una fracción:

$$\text{Fracción} = \frac{\text{numerador}}{\text{denominador}} = \frac{\text{parte}}{\text{entero}}$$

Al calcular una proporción:

$$\text{Proporción} = p \text{ [del grupo total en una categoría]} = \frac{\# \text{ en una categoría}}{\# \text{ en el grupo total}} = \text{cociente}$$

Al calcular un porcentaje:

$$\text{Porcentaje} = \% \text{ [del grupo total en una categoría]} = p (100)$$

Al calcular una tasa:

$$\text{Tasa de ocurrencia} = (p) \text{ (un número base útil)}$$

Al calcular un cambio de porcentaje:

$$\text{Cambio de porcentaje} = \frac{\# \text{ al tiempo 2} - \# \text{ al tiempo 1}}{\# \text{ al tiempo 1}}$$

Preguntas para el capítulo 1

1. Un entrevistador en un estudio entendió mal a un encuestado y le anotó incorrectamente la edad. ¿Fue ésta una equivocación o un error estadístico? Explique.
2. Mary Jones se ha preocupado por los desastres naturales, al notar que en un solo año hubo inundaciones en el medio oeste, sequía en el sur y grandes terremotos en el oeste. Ella cree que dichos eventos constituyen la prueba de que el

fin del mundo está cerca. Considerando que Mary tiene una imaginación vivida, explique por qué carece de imaginación estadística.

3. En un estudio sobre los *estudiantes del último año* en una prestigiada universidad, se mide el área de especialización de los alumnos (psicología, sociología, química, inglés, arte, etcétera) y su año de escolaridad (primero, segundo, tercero, último año). En dicho estudio, ¿cuál de estas mediciones representa una variable, y cuál una constante?
4. En un estudio sobre los estudiantes universitarios del último año, se mide su promedio académico y su consumo de alcohol durante el mes anterior. Formule una hipótesis para estas dos variables e indique cuál es la variable independiente y cuál es la variable dependiente. ¿Podrían generalizarse los resultados de *estudiantes del último año al conjunto de estudiantes* de la universidad? ¿Por qué?
5. Para una muestra de personas sin hogar, usted se interesa en la relación entre género y tipos de lugares para dormir (donde el sujeto pasó la noche anterior). ¿Cuál es la variable independiente, y cuál la dependiente?
6. Bob posee una tienda de libros y computadoras llamada InfoManiacs. Él calculó la proporción de sus ganancias que resultan de vender programas de computadora y obtuvo una respuesta de 2.49. ¿Esto podría ser correcto? Explique su respuesta.
7. ¿Cuál es la característica esencial de la ciencia que la distingue de otras formas de indagar sobre la naturaleza?
8. Identifique una creencia común que usted sospeche que se trate de un mito. Sugiera qué tipos de datos serían recolectados para descubrir el mito. ¿Cómo se podría aplicar el pensamiento proporcional para desafiar esta falsedad ampliamente sostenida?
9. Para darse una idea de qué tan organizados están los procedimientos científicos, vaya a una hemeroteca y hojee varias publicaciones científicas, como *American Sociological Review*, *American Journal of Sociology*, *Journal of the American Psychological Association*, *Journal of Health and Social Behavior*, *Administrative Science Quarterly*, *Criminology*, *Social Services Review*, *American Journal of Psychology*, *American Political Science Review*, *Review of Public Administration* y *Political Science Quarterly*. Note la abundancia de tablas estadísticas y gráficas en esos artículos. Observe también que todos los artículos en los diferentes volúmenes tienen encabezados de secciones similares.
 - a) Enliste los títulos de por lo menos cinco artículos de tres diferentes revistas.
 - b) Compare tales listas con las siete fases del proceso de investigación y comente el resultado.
10. Suponga que, en Estados Unidos, un estado tuvo una tasa de mortalidad infantil de 8.6 muertes por 1 000 nacidos vivos en 1998. En el año 2000, el departamento de salud pública estatal organiza una conferencia llamada *Metas 2010*, donde políticos y funcionarios del gobierno buscan mejorar la salud pública en el nuevo milenio. Ellos establecieron una tasa óptima de 6.0 muertes infantiles por 1 000 nacidos vivos para el año 2010. Esta tasa óptima es un

11. En la mayoría de los estados, el límite de velocidad interestatal es de 70 millas por hora. Se toman muestras aleatorias de la velocidad de los vehículos, con un radar detector de velocidad. Se determina que la velocidad promedio es de 74 millas por hora. El límite de velocidad establecido es un _____ estadístico, mientras que 74 millas por hora es un _____ estadístico.

Ejercicios para el capítulo 1

1. Complete los espacios en blanco de la siguiente tabla (consulte el apéndice A como repaso):

	Fracción	Proporción	Porcentaje
a)	$\frac{86}{211}$.4076	_____
b)	$\frac{1}{656}$	_____	_____
c)	$\frac{28}{10\ 000}$.0028	_____
d)	_____	.1711	_____
e)	_____	_____	23.22%
f)	_____	_____	89.20%

2. Complete los espacios en blanco de la siguiente tabla (véase el apéndice A para repaso):

	Fracción	Proporción	Porcentaje
a)	$\frac{39}{192}$.2031	_____
b)	$\frac{18}{423}$	_____	_____
c)	$\frac{441}{10\ 000}$.0441	_____
d)	_____	.0877	_____
e)	_____	_____	45.50%

3. En una conversación casual después de clase, Jane y Ana descubren que poseen un hábito en común. Para liberar tensión, ellas observan qué tan frecuentemente pueden atinarle a un cesto de basura lanzando una pelota de papel. Jane hace alarde de que en 250 tiros, acertó 128 veces. En 265 tiros, Ana "encestó" 157. ¿Quién es la mejor lanzadora? ¿Por qué?
4. Como estadístico de una liga de basquetbol de una preparatoria local, usted está encargado de llevar las estadísticas de los puntos anotados en los juegos. Hasta ahora, Antonio "Tiro Caliente" Johnson ha intentado 73 tiros y ha acertado 34, mientras Clarence "Golpeador" Williams ha intentado 52 tiros y ha enceestado 28. ¿Qué jugador tiene el porcentaje de anotación más alto?
5. Para cumplir con las recomendaciones de la Secretaría de Protección del Ambiente, la cantidad de partículas suspendidas en el aire de la ciudad puede exceder 69 partes por millón, sólo el 15% de los días del año sin que se tomen medidas precautorias. ¿Cuántos días sucede esto?
6. Scott, Sam y Sid, tres amigos que empacan comestibles en un mercado local, decidieron juntar sus propinas una tarde para comprar un regalo para su amiga Cindy, quien convalecía en el hospital. Scott cooperó con \$15, Sam \$12 y Sid \$10. ¿Con qué proporción de dinero contribuyó cada uno para el regalo?
7. Un balneario va a seleccionar a 100 miembros para asistir a las Olimpiadas de invierno, mitad hombres y mitad mujeres. Los grupos de edad serán representados según la membresía proporcional. En otras palabras, si la membresía se compone en su mayoría de 31 a 40 años, principalmente individuos de este grupo de edad harán el viaje. La siguiente tabla muestra la composición de la membresía por edad y género. Complete las columnas de proporción (*p*) y número (# de asistentes) para cada grupo de edad por género. Muestre la fórmula general y un ejemplo de los cálculos para al menos un grupo de edad.

Grupo de edad	Hombres	<i>p</i>	# de asistentes	Mujeres	<i>p</i>	# de asistentes
21-30	49			80		
31-40	170			217		
41-50	169			176		
51-60	84			91		
61+	22			48		
Totales	494		50	612		50

8. Una corporación con oficinas en Los Ángeles y Nueva York capacitará a sus asistentes administrativos en un nuevo programa de cómputo. Cada sesión integrará a 40 personas. Para que empleados de cada departamento reciban capacitación de inmediato, los participantes son elegidos utilizando cuotas por departamento y ciudad. Por ejemplo, si 10 por ciento de los asistentes administrativos están en el departamento de mercadotecnia en Los Ángeles, 10 por ciento de los aprendices vendrán de ese departamento. La siguiente tabla muestra la composición por departamento y ciudad. Complete las columnas de proporción (*p*) y número (# de asistentes) para cada departamento y ciudad. Escriba

ba la fórmula general y, por lo menos, dé un ejemplo de los cálculos para una categoría de departamento-ciudad.

Departamento (sección)	Los Ángeles	p	# de asistentes	Nueva York	p	# de asistentes
Personal	36			43		
Mercadotecnia	81			93		
Envíos	65			78		
Administración	24			31		
Contabilidad	25			38		
Totales	231		40	283		40

9. Usted está interesado en el fenómeno de las placas de matrícula personalizadas, aquellas donde el dueño de un vehículo lleva su nombre o un dicho. En una muestra aleatoria de 341 placas, usted encuentra que 73 están personalizadas. ¿Qué proporción de placas no lo están?
10. Hay 768 ranas en un estanque: 61 por ciento son machos, 75 por ciento del total tienen verrugas. Suponiendo que machos y hembras tienen la misma probabilidad de estar "averrugadas", ¿cuántas ranas macho esperaría que tuvieran verrugas?
11. Complete la siguiente tabla calculando la tasa de confinamiento en una institución (prisiones y hospitales para enfermos mentales) por 100 mil habitantes. Exponga la fórmula general y los cálculos para Anderson, Indiana.

Ciudad	Población	Número de personas confinadas en una institución	Tasa por 100 000 habitantes
Anderson, Indiana	130 669	3 981	
Bellingham, Washington	127 780	1 602	
Duluth, Minnesota	239 971	4 610	
Modesto, California	370 522	4 456	

12. Complete la siguiente tabla calculando la tasa de confinamiento en una institución (prisiones y hospitales mentales) por 100 mil habitantes. Muestre la fórmula general y los cálculos para Bakersfield, California.

Ciudad	Población	Número de personas confinadas en una institución	Tasa por 100 000 habitantes
Bakersfield, California	543 477	10 808	
Burlington, Carolina del Norte	108 213	1 158	
Great Falls, Montana	77 691	787	
Poughkeepsie, Nueva York	259 462	11 082	

13. Turner (1995) investigó los efectos del desempleo. Contactó a 5 612 personas, a quienes consideró elegibles para el estudio porque habían estado desempleadas por lo menos una vez, desde que se incorporaron a la fuerza laboral. De estas personas elegibles, realmente entrevistó a 3 617, entre las cuales, 1 252 se integraron en un estudio a largo plazo. En el grupo de estudio a largo plazo, 154 estuvieron "recientemente desempleadas", pues perdieron sus trabajos en los últimos tres años; de ellas, 45 permanecían desempleadas.
- ¿Qué proporción de los sujetos elegibles fue entrevistada?
 - ¿Qué porcentaje de los que fueron entrevistados *no* participaron en el estudio a largo plazo?
 - ¿De los que sí participaron, qué porcentaje quedó desempleado en los últimos tres años?
 - ¿Qué porcentaje de los recientemente desempleados regresó a trabajar?
14. Turner (1995) investigó los efectos del desempleo. Él contactó a 5 612 personas, quienes consideró elegibles para el estudio porque habían estado desempleadas por lo menos una vez, desde que se incorporaron a la fuerza laboral. De estas personas elegibles, realmente se entrevistó a 3 617, entre las cuales 1 252 se identificaron para un estudio a largo plazo. Imagine que de las 1 252 personas en el grupo de estudio a largo plazo, 732 eran hombres y 520 eran mujeres. Entre las 154 recientemente desempleadas, 80 eran hombres y 74 mujeres. Entre las 45 que aún permanecían desempleadas, 25 eran hombres y 20 mujeres.
- En este grupo de estudio a largo plazo, ¿qué género tuvo mayor proporción de desempleo reciente?
 - Entre los recientemente desempleados, ¿fueron hombres o mujeres los más afortunados al regresar a la fuerza de trabajo?
15. Es el regreso del fin de semana y en el campus se eligió a un rey y una reina. Cuando se anunciaron los resultados para la coronación de la reina, se supo que Natalia ganó con 526 votos; la ganadora del segundo lugar obtuvo 510; las siguientes cuatro recibieron 482, 325, 200 y 150 votos, respectivamente, y las restantes 10 participantes alcanzaron un total de 1 140 votos. En la fiesta para los concursantes, usted descubre que Natalia viene de una familia numerosa, y que 28 de sus parientes asisten a la misma universidad y votaron por ella.
- ¿Qué porcentaje de los 15 000 estudiantes votó en la elección?
 - ¿Del total de votos, qué porcentaje recibió Natalia?
 - ¿Qué proporción del total de estudiantes representa la familia de Natalia? Simplifique esta respuesta calculando una tasa.
 - ¿Fue el apoyo familiar un factor clave en la victoria de Natalia?
16. En un estudio internacional de afiliación religiosa, un investigador obtuvo los siguientes datos de 2 465 encuestados seleccionados aleatoriamente (datos ficticios). Complete la tabla en las columnas de proporción (p) y porcentaje (%) de encuestados para cada religión.

Religión	Número	<i>p</i>	%
Cristianos	814		
Musulmanes	444		
Hindúes	320		
Budistas	148		
Otras religiones	345		
Personas sin religión	394		
Totales	2 465		

17. Para la Oficina de Censos de Estados Unidos, un Área Estadística Metropolitana (AEM) es un condado con una ciudad central de 50 000 o más habitantes junto con los condados adyacentes, cuyas economías están estrechamente ligadas con la ciudad central. Los siguientes datos del censo de 1990 en Estados Unidos presentan poblaciones de las AEM clasificadas por su situación residencial.
- Complete la columna de "Totales" en la tabla.
 - Calcule la proporción (*p*) para cada área residencial en cada AEM. Exponga la fórmula general y un ejemplo de los cálculos para al menos uno de los casos.
 - ¿Qué AEM parece más rural que las otras?

Área Estadística Metropolitana 1 (AEM)	Distribución residencial						Totales
	Rural urbano	<i>p</i>	Rural no agrícola	<i>p</i>	Agrícola	<i>p</i>	
Abilene, TX,	107 052		11 620		983		
Bakersfield, CA	455 300		84 870		3 307		
Bellingham, WA	75 697		48 487		3 596		
Duluth, MN	167 424		71 271		1 276		
Poughkeepsie, NY	146 526		111 748		1 188		

18. Para la Oficina de Censos de Estados Unidos, un Área Estadística Metropolitana (AEM) es un condado con una ciudad central de 50 000 o más habitantes junto con los condados adyacentes, cuyas economías están estrechamente ligadas con la ciudad central. Los siguientes datos del censo de 1990 en Estados Unidos presentan poblaciones EAM de las clasificadas por la situación residencial dentro de la AEM.
- Complete la columna de "Totales" en la tabla.
 - Calcule la proporción (*p*) para cada área residencial en cada AEM. Muestre la fórmula general y un ejemplo de los cálculos para al menos uno de los casos.
 - ¿Qué AEM parece más rural que las otras?

Área Estadística Metropolitana 1 (AEM)	Distribución residencial						Totales
	Rural urbano	<i>p</i>	Rural no agrícola	<i>p</i>	Agrícola	<i>p</i>	
Abilene, TX	107 052		11 620		983		
Anderson, IN	87 438		40 634		2 597		
Battle Creek, MI	95 188		37 967		2 627		
Burlington, NC	71 289		35 439		1 485		
Great Falls, NY	63 531		12 341		1 819		

19/20. Resuelva esta vieja adivinanza.

Cuando iba a St. Ives, encontré a un hombre con siete esposas.
Cada esposa tenía siete sacos, cada saco tenía siete gatos.
Cada gato tenía siete gatitos.
Gatitos, gatos, sacos y esposas, ¿cuántos iban a St. Ives?

Aplicaciones opcionales en computadora para el capítulo 1

Si en su clase emplea las aplicaciones para computadora optativas que acompañan este texto, siga las instrucciones básicas en el disco compacto de *Computer Applications for The Statistical Imagination*, para acceder a la información allí contenida. No se requiere ninguna experiencia anterior con computadoras. El disco incluye el programa *Statistical Package for the Social Sciences (SPSS), for Windows, Student Version*. Además, hay una variedad de conjuntos de datos, con descripciones completas, ejercicios del capítulo relacionados con estos conjuntos de datos e ilustraciones detalladas de cómo interpretar la información de la computadora. Pueden usarse versiones anteriores de SPSS; pero la información resultante aparecerá con un formato diferente.

El SPSS for Windows, Student Version es más que adecuado para aprender estadísticas básicas. Si usted quiere dirigir su propia investigación, sin embargo, quizá desee el acceso al sistema base completo de la versión regular (no del estudiante), del SPSS for Windows o el SPSS for Windows Graduate Pack. Estas versiones del SPSS tienen varias ventajas, incluso una ventana de "Syntax Editor" que pega y guarda comandos seleccionados con el ratón, para su uso posterior; no existen limitaciones en el número de variables o el tamaño de las muestras en los archivos de datos.

Los conjuntos de datos en el disco compacto *Computer Applications for The Statistical Imagination* se han modificado un poco para facilitar su comprensión y, por consiguiente, no satisfacen los propósitos de la investigación real. Usted puede solicitar el conjunto original de datos sin modificar las fuentes mencionadas en el disco. Otros datos están disponibles en Internet por agencias gubernamentales y fundaciones de investigación como la Inter University Consortium for Political and Social Research (ICPSR) y la National Opinion Research Center (NORC).

Para acceder a los ejercicios del capítulo 1, inserte el disco compacto en el drive del CD ROM, haga clic en "Chapter Exercises" y luego en "Chapter 1". Este primer

ejercicio incluye orientación para el software estadístico *SPSS for Windows* con instrucciones sobre cómo recuperar archivos del disco compacto. Una vez que el *SPSS for Windows* se activa, haga clic en *Help* y luego en *Tutorial*. Siga el tutorial para que se familiarice con las ventanas, iconos y menús básicos.

Notas

1. Cuando alguien realiza una declaración objetiva sobre un objeto (o persona o situación), la declaración describe una característica que es realmente parte del objeto, por ejemplo, la declaración: "La señal de alto es roja." Cuando alguien hace una declaración subjetiva, la declaración describe una característica del "sujeto" observador, más que del objeto. Las declaraciones subjetivas constituyen, por consiguiente, ideas personales u opiniones que reflejan los prejuicios, distorsiones, puntos de vista personales o deformaciones del individuo que hace la declaración. Por ejemplo, alguien que es daltónico diría: "La señal de alto es gris." Lo "grisáceo" no es parte de la señal, es sólo la percepción del sujeto observador.
2. Este dato fue proporcionado por Jerry Tracey, meteorólogo de televisión de WVTM, Birmingham, Alabama.
3. La abreviación latina *i.e.* significa "esto es" (*id est*); *e.g.* significa "por ejemplo" (*exempli gratia*).
4. El término *independiente* viene de la ciencia del laboratorio, donde variables predictoras se manipulan independientemente de los resultados. Por ejemplo, en un estudio sobre los efectos de una droga en ratas, la droga se administra a algunas ratas (el grupo experimental); mientras un placebo (o droga falsa) se da a otro grupo igual (el grupo control). La decisión de qué ratas son asignadas a cada grupo, se hace independientemente de medir qué ratas mejoran.
5. El término *teoría* a menudo se utiliza para designar una idea no corroborada, como en "esto es sólo una teoría"; tal sentido no es el que se le atribuye en la ciencia. Las teorías científicas *no* están basadas en conjeturas u opiniones, sino en el análisis objetivo de datos cuidadosamente recolectados.

CAPÍTULO

2

ANÁLISIS ESTADÍSTICO: MANEJO Y CONTROL DEL ERROR

Introducción	34
Control del error de muestreo	35
Estimación estadística cuidadosa contra adivinación o estimación apresurada	38
Error de muestreo y su manejo con la teoría de la probabilidad.	39
Control del error de medición	41
Niveles de medición: selección cuidadosa de los procedimientos estadísticos	42
Medición	42
Variables nominales	43
Variables ordinales	44
Variables de intervalo	44
Variables de razón	46
Codificación y conteo de las observaciones	47
Distribuciones de frecuencias	51
Estandarización de distribuciones de puntuaciones	51
Comparación de las frecuencias de dos variables nominales/ordinales	53