INVESTIGACIÓN DEL COMPORTAMIENTO

Cuarta edición

FRED N. KERLINGER (†)

HOWARD B. LEE
California State University

Traducción
Leticia Esther Pineda Ayala
Ignacio Mora Magaña
Traductores profesionales

Revisión técnica
Mtra. Cecilia Balbás Diez Barroso
Coordinadora del Área de Psicología Educativa
Escuela de Psicología
Universidad Anáhuac

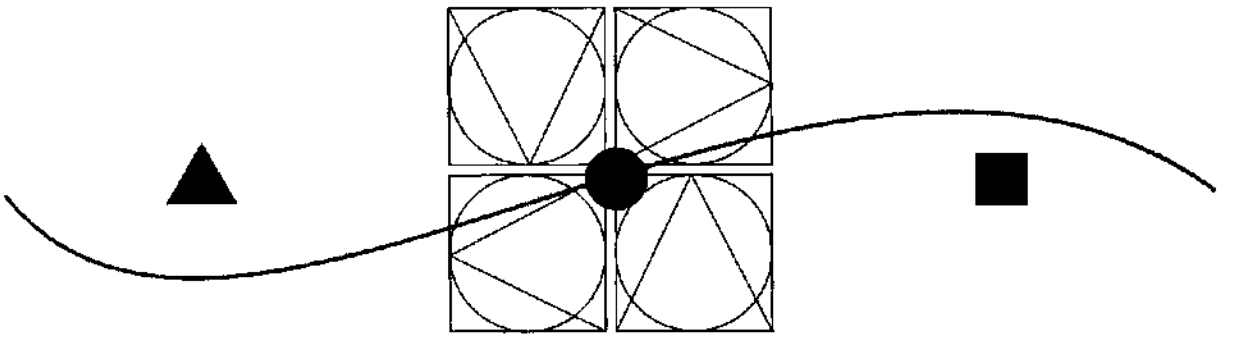
Dra. Guadalupe Vadillo Bueno
Psicóloga y Master en Educación
Universidad de las Américas
Doctora en Educación
Universidad La Salle
Directora de Educación Continua y de Comunicación Humana
Universidad de las Américas

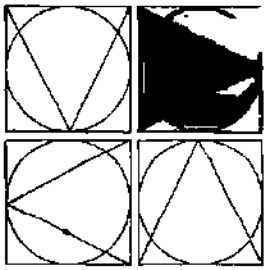
McGRAW-HILL
MÉXICO • BUENOS AIRES • CARACAS • GUATEMALA • LISBOA • MADRID
NUEVA YORK • SAN JUAN • SANTAFÉ DE BOGOTÁ • SANTIAGO • SÃO PAULO
AUCKLAND • LONDRES • MILÁN • MONTREAL • NUEVA DELHI
SAN FRANCISCO • SINGAPUR • ST. LOUIS • SIDNEY • TORONTO

Contenido

Parte Uno

El lenguaje y enfoque de la ciencia





Capítulo 1

La ciencia y el enfoque científico

- Ciencia y sentido común
- Cuatro métodos del conocimiento
- La ciencia y sus funciones
- Los objetivos de la ciencia, explicación científica y teoría
- La investigación científica: definición
- El enfoque científico
  - **Problema-obstáculo-idea**
  - **Hipótesis**
  - **Razonamiento-deducción**
  - **Observación-prueba-experimento**

Para entender cualquier actividad humana compleja es necesario comprender el lenguaje y el enfoque de quienes la realizan. Así sucede con la ciencia y la investigación científica. Se debe conocer y entender, al menos en parte, el lenguaje científico y el enfoque científico en la solución de problemas.

Una de las cosas que más confunde al estudiante de ciencia es la forma particular en que los científicos utilizan palabras ordinarias y, para complicar el asunto, inventan incluso nuevas palabras. Hay buenas razones para este uso tan especializado del lenguaje que serán evidentes más adelante. Por ahora basta decir que es necesario entender y aprender el lenguaje que usan los científicos sociales. Cuando los investigadores hablan de sus variables dependientes e independientes, se debe saber a qué se refieren. Cuando dicen que han aleatorizado sus procedimientos experimentales no sólo es necesario saber a qué se refieren, sino entender por qué lo hacen.

De igual forma, la manera en que el científico se aproxima a los problemas debe ser entendido con claridad. No porque su enfoque sea muy diferente al de cualquier persona. Por supuesto que sí *es* distinto, pero no es extraño ni esotérico. Por el contrario, cuando se

comprende la labor del científico parece natural y casi inevitable. De hecho, es probable que nos preguntemos por qué una gran parte del pensamiento humano y de la solución de problemas no están tan conscientemente estructurados de la misma manera.

El propósito de los capítulos 1 y 2 de este libro es ayudar al estudiante a aprender y entender tanto el lenguaje como el enfoque de la ciencia y de la investigación. En estos capítulos se estudiarán muchos de los constructos básicos del investigador social, conductual y educativo. En algunos casos no será posible dar definiciones completas y satisfactorias debido a la falta de antecedentes en este punto tan temprano del desarrollo del lector. En tales casos se intentará formular y usar un primer enfoque razonablemente preciso, para de ahí progresar a definiciones más satisfactorias. Comencemos nuestro estudio considerando cómo aborda el científico los problemas y cómo este enfoque difiere de lo que puede llamarse un enfoque por sentido común.

Ciencia y sentido común

Al inicio del siglo XX, Whitehead (1911/1992, p. 157) señaló que en el pensamiento creativo el sentido común es un mal instructor. "Su único criterio para juzgar es que las ideas nuevas deben parecerse a las viejas." Tiene razón: el sentido común puede a menudo ser un mal consejero para evaluar el conocimiento. ¿Pero, en qué se parecen y en qué difieren la ciencia y el sentido común? Desde un punto de vista ambos se asemejan: sus defensores dirían que la ciencia es una extensión sistemática y controlada del sentido común. James Bryant Conant (1951) establece que el sentido común es una serie de conceptos y esquemas conceptuales[1] satisfactorios para los usos prácticos de la humanidad. Sin embargo, estos conceptos y esquemas conceptuales pueden ser engañosos en la ciencia moderna —en particular en psicología y educación—. Muchos educadores del siglo XIX, por sentido común usaban el castigo como herramienta básica de la pedagogía. Sin embargo, a mediados del siglo XX la evidencia mostró que esta perspectiva de la motivación basada en el sentido común podía ser bastante errónea. La recompensa parece ser más efectiva que el castigo para apoyar el aprendizaje. Sin embargo, recientes hallazgos sugieren que diferentes formas de castigo son útiles en el aprendizaje en el salón de clases (Marlow *et al.*, 1997; Tingstrom *et al.*, 1997). La ciencia y el sentido común difieren marcadamente en cinco aspectos. Estos desacuerdos giran alrededor de las palabras *sistemático* y *controlado*.

Primero, los usos de los esquemas conceptuales y de las estructuras teóricas son notablemente diferentes. La persona común puede usar "teorías" y conceptos, pero en general lo hace de una forma vaga y a menudo acepta sin reparo explicaciones fantásticas de los fenómenos humanos y naturales. Por ejemplo, puede creerse que una enfermedad es un castigo por haber pecado (Klonoff y Landrine, 1994); o que se es alegre porque se tiene sobrepeso. Los científicos, por otro lado, construyen estructuras teóricas de forma sistemática, luego evalúan su consistencia interna, y someten algunos de sus aspectos a una

[1] Un *concepto* es una palabra que expresa una abstracción formada por la generalización de elementos particulares: "agresión" es un concepto, una abstracción que expresa un número de acciones particulares que tienen la característica común de dañar personas u objetos. Un *esquema conceptual* es un conjunto de conceptos interrelacionados por proposiciones hipotéticas y teóricas. Un *constructo* es un concepto con el significado adicional de haber sido creado o adaptado para propósitos científicos especiales. "Masa", "energía", "hostilidad", "introversión" y "rendimiento" son constructos que pueden ser llamados con más precisión "tipos construidos"; o "clases construidas", las clases o grupos de objetos o eventos se agrupan porque poseen una característica común definida por el científico. En un capítulo posterior se definirá el término "variable". Por ahora es suficiente con saber que representa un símbolo o nombre de una característica que puede adoptar diferentes valores numéricos.

prueba empírica. Además, se percatan de que los conceptos que emplean son términos acuñados por el hombre que pueden o no mostrar una relación estrecha con la realidad.

Segundo, los científicos prueban sus teorías e hipótesis de forma sistemática y empírica. Quienes no son científicos también prueban "hipótesis", pero lo hacen de manera selectiva: con frecuencia "seleccionan" evidencia sólo porque es consistente con las hipótesis. Veamos un estereotipo: los asiáticos tienen vocación científica y matemática. Si la gente lo cree podrán "verificar" con facilidad esta creencia al notar que muchos asiáticos son ingenieros y científicos (véase Tang, 1993). No se perciben las excepciones al estereotipo: los asiáticos que no son científicos o no son matemáticos. Los científicos sociales y conductuales identifican estas "tendencias de selección" como fenómeno psicológico común, y se cuidan mucho de no contaminar su investigación con sus propias preconcepciones o predilecciones y con el apoyo selectivo de hipótesis. Por principio, no se sienten satisfechos con realizar una exploración somera de las relaciones; los científicos deben probar estas relaciones en el laboratorio o en el campo. Ellos no se contentan por ejemplo, con las supuestas relaciones entre métodos de enseñanza y rendimiento, entre inteligencia y creatividad, entre valores y decisiones administrativas. Insisten en probar estas relaciones de manera sistemática, controlada y empírica.

Una tercera diferencia yace en la noción de control. En la investigación científica control significa diversas cosas. Por ahora considere que el científico trata sistemáticamente de descartar las variables que son posibles "causas" de los efectos bajo estudio, de otras variables que se ha hipotetizado son las "causas". La gente común rara vez se preocupa por controlar sus explicaciones de los fenómenos observados sistemáticamente. Por lo general, poco se esmeran por controlar las fuentes extrañas de influencia, y tienden a aceptar más aquellas explicaciones que concuerdan con sus preconcepciones y sesgos. Si creen que los barrios bajos producen delincuencia, tienden a hacer caso omiso de la delincuencia en otras zonas. Los científicos, por otro lado, buscan y "controlan" la incidencia de la delincuencia en diferentes tipos de barrios. La diferencia, por supuesto, es profunda.

Otra distinción entre ciencia y sentido común quizás no es tan clara. Antes se dijo que el científico está preocupado de manera constante por las relaciones entre fenómenos. La persona común también, pero usa su sentido común para explicar los fenómenos. El científico, sin embargo, persigue las relaciones de forma sistemática y concienzuda. La manera en que la persona común pretende conocer estas relaciones es relajada, no sistemática, y sin control: por ejemplo, con frecuencia observa la ocurrencia fortuita de dos fenómenos y los liga de inmediato de forma indisoluble como causa y efecto.

Tomemos la relación probada en un estudio clásico realizado hace muchos años por Hurlock (1925). Usando terminología más reciente, esta relación se puede expresar como: el reforzamiento positivo (recompensa) produce un mayor incremento de aprendizaje que el castigo. La relación se encuentra entre el reforzamiento (o recompensa y castigo) y el aprendizaje. Los educadores y padres de familia del siglo XIX con frecuencia suponían que el castigo era el agente más efectivo en el aprendizaje. Los educadores y padres de hoy asumen constantemente que el reforzamiento positivo (recompensa) es más efectivo. En ambos casos podemos decir que el punto de vista se basa sólo en el "sentido común". Es obvio, podrían decir, que si se premia o castiga a un niño aprenderá mejor. Por otro lado, el científico, ya sea que defienda en lo personal uno, otro o ninguno de estos puntos de vista, probablemente insistirá en una prueba sistemática y controlada de ambas relaciones (y de otras), como lo hizo Hurlock. Al usar el método científico Hurlock encontró que los incentivos estaban estrechamente relacionados con el rendimiento en aritmética. El grupo que fue elogiado obtuvo mayores puntuaciones que los que fueron reprobados o ignorados.

Una última diferencia entre el sentido común y la ciencia estriba en las diferentes explicaciones de los fenómenos observados. Cuando el científico intenta explicar las rela-

ciones entre los fenómenos observados descarta cuidadosamente lo que se ha llamado "explicaciones metafísicas". Una explicación metafísica es simplemente una proposición que no puede ser probada. Decir por ejemplo que la gente es pobre y padece hambre porque Dios así lo dispuso, o decir que es malo ser autoritario, es hablar metafísicamente.

Ninguna de estas proposiciones se puede probar; por lo tanto, son metafísicas y no son del interés de la ciencia. Esto no significa que los científicos necesariamente desprecien tales afirmaciones, digan que no son ciertas o mantengan que carecen de sentido. Sólo implica que *como científicos* no les interesan. En pocas palabras, la ciencia está involucrada con asuntos que pueden observarse y probarse. Si las propuestas o preguntas no implican esta observación y prueba públicas, no constituyen propuestas o preguntas científicas.

Cuatro métodos del conocimiento

Charles Sanders Peirce, como indica Buchler (1955), señaló cuatro formas generales de conocer o, como lo indicó, de establecer creencias. En la siguiente discusión los autores se tomaron algunas libertades con la propuesta original de Peirce en un intento de aclarar las ideas y hacerlas más apropiadas a esta presentación. El primero es el *método de la tenacidad*. En él la gente sostiene firmemente la verdad, la cual asumen como cierta debido a su apego a ella y a que siempre la han considerado como verdadera y real. La frecuente repetición de tales "verdades" parece aumentar su validez. A menudo la gente se aferra a sus creencias aun frente a hechos que claramente están en conflicto con ellas. Además infieren "nuevo" conocimiento a partir de proposiciones que pueden ser falsas.

Un segundo método de conocimiento o de establecer creencias es el *método de la autoridad*, o de creencias establecidas. Si la Biblia lo dice así es. Si un notable físico dice que hay un Dios, lo hay. Si una idea cuenta con el peso de la tradición y la sanción pública para apoyarla, entonces así. Peirce señaló que este método es superior al de la tenacidad porque es posible lograr progreso humano, aunque de manera lenta. De hecho, la vida no podría funcionar sin el método de la autoridad. Dawes (1994) establece que, como individuos, no podemos saberlo todo. Los norteamericanos reconocen la autoridad de la Oficina de Administración de Drogas y Alimentos de Estados Unidos (FDA) para determinar que cuanto comen y beben es seguro. Dawes estableció que no existe la mente completamente abierta que cuestione toda autoridad. Debemos asumir una gran cantidad de hechos e información con base en la autoridad. Por lo tanto, no podemos concluir que el método de la autoridad sea defectuoso; lo es sólo bajo ciertas circunstancias.

El *método a priori* es la tercera forma de conocimiento o de establecer creencias. Graziano y Raulin (1993) lo llamaron el *método de la intuición*. Basa su superioridad en el supuesto de que las proposiciones aceptadas por el "a priorista" son por sí mismas evidentes. Observe que la proposición *a priori* "concuerda con la razón" y no necesariamente con la experiencia. La idea parece ser que la gente, a través de la comunicación y trato libres pueda alcanzar la verdad porque sus inclinaciones naturales tienden hacia ella. La dificultad con esta postura subyace en la expresión "concuerda con la razón". ¿La razón de quién? Imagine a dos sujetos honestos y bien intencionados que usando el proceso racional alcanzan diferentes conclusiones. ¿Quién está en lo correcto? ¿Es cuestión de gustos, como lo señaló Peirce? Si algo es patente para muchas personas —por ejemplo que el aprender materias difíciles entrena la mente y fortalece el carácter moral, que la educación estadounidense es inferior a la asiática y europea— ¿significa que en realidad lo sea? De acuerdo con el método *a priori* lo es, justamente porque se mantiene frente a la razón.

El cuarto método es el *método de la ciencia*. Dice Peirce:

> Para satisfacer nuestras dudas..., por lo tanto, es necesario encontrar un método por el que nuestras creencias se determinen no a partir de algo humano, sino por algo con permanencia externa, por algo que nuestro pensamiento no pudiera afectar ... El método debe ser tal que la conclusión última de todo hombre fuera la misma. Éste es el método de la ciencia. Su hipótesis fundamental es ésta: "hay cosas reales cuyas características son totalmente independientes de nuestra opinión acerca de ellas..." (Buchler, 1995, p. 18).

El enfoque científico tiene una característica de la que carecen los otros métodos de obtención del conocimiento: autocorrección. Hay puntos de verificación intrínsecos a lo largo de todo el camino del conocimiento científico. Estos controles están concebidos y se utilizan de manera tal que dirigen y verifican las actividades científicas y las conclusiones con el fin de obtener conocimiento del que se pueda depender. Incluso si una hipótesis parece sustentarse en un experimento, el científico probará hipótesis alternas posibles que, si también reciben apoyo, pueden generar dudas sobre la primera hipótesis. Los científicos no aceptan declaraciones como verdaderas aunque en principio la evidencia pueda parecer prometedora. Insisten en probarlas. También subrayan la necesidad de que cualquier procedimiento de prueba esté abierto al escrutinio *público*. Una interpretación del método científico es que no hay método científico específico. Más bien existe una variedad de métodos que el científico puede emplear y, de hecho, usa, pero probablemente pueda decirse que hay un solo enfoque científico.

Como dijo Peirce, los controles usados en la investigación científica están anclados tanto como es posible en la realidad, más allá de las creencias personales del científico, y de sus percepciones, sesgos, valores, actitudes y emociones. Tal vez la mejor palabra para expresar esto es *objetividad*. La objetividad constituye el acuerdo entre jueces "expertos" sobre lo observado o sobre lo que se hizo o hará en la investigación (véase Kerlinger, 1979, para un análisis de objetividad, su significado y su carácter controvertido). De acuerdo con Sampson (1991, p. 12) la objetividad "se refiere a aquellas declaraciones acerca del mundo que se pueden justificar y defender en el presente usando los estándares de argumento y prueba empleados en la comunidad a la que pertenecemos —por ejemplo, la de científicos"—. Pero, como se verá más adelante, el enfoque científico involucra más que estas dos declaraciones. El hecho es que alcanzamos un conocimiento del que podemos depender más debido a que la ciencia apela finalmente a la evidencia: las proposiciones se someten a la prueba empírica. Puede surgir otra objeción: la teoría, que los científicos usan y exaltan, proviene de gente, de los científicos mismos. Pero, como Polanyi (1958/1974 p. 4) señala, "una teoría es algo distinto a mí". Así, una teoría ayuda al científico a lograr una mayor objetividad. En pocas palabras, los científicos sistemática y conscientemente usan los aspectos autocorrectivos del enfoque científico.

La ciencia y sus funciones

¿Qué es la ciencia? Esta pregunta no es fácil de contestar; de hecho, no intentaremos presentar definición alguna de ciencia de manera directa. En cambio hablaremos de nociones y perspectivas de la ciencia para después intentar explicar sus funciones.

Ciencia es una palabra malinterpretada. Parece ser que hay tres estereotipos populares que dificultan el entendimiento de la actividad científica. Uno es el de bata blanca-estetoscopio-laboratorio. Se percibe a los científicos como individuos que trabajan con hechos en laboratorios; usan equipo complicado, hacen muchos experimentos y amontonan hechos con el propósito final de perfeccionar a la humanidad. Así, aunque sean exploradores

poco imaginativos en busca de hechos, se les redime por sus nobles motivos. Puede creérseles cuando, por ejemplo, dicen que tal o cuál dentífrico es bueno para usted, o que no debería fumar cigarrillos.

El segundo estereotipo de los científicos consiste en que son individuos brillantes que piensan, elaboran teorías complejas y pasan el tiempo en torres de marfil alejados del mundo y sus problemas. Son teóricos poco prácticos, aun cuando su pensamiento y teorías ocasionalmente tengan resultados de significación práctica, como la energía atómica.

El tercer estereotipo equipara erróneamente a la ciencia con la ingeniería y la tecnología: la construcción de puentes, el mejoramiento de automóviles y misiles, la automatización de la industria, la invención de máquinas para enseñar. El trabajo del científico, según este estereotipo, está dedicado a optimizar inventos y artefactos. Se concibe al científico como una clase de ingeniero altamente especializado que trabaja para hacer la vida más cómoda y eficiente.

Estos estereotipos limitan al estudiante para entender la ciencia, las actividades y el pensamiento del científico, y la investigación científica en general. En pocas palabras, hacen que la tarea del alumno sea más difícil de lo que podría resultar. Por ello, se deben eliminar para hacer espacio a nociones más apropiadas.

Hay dos amplias visiones de la ciencia: la estática y la dinámica. De acuerdo con Conant (1951, pp. 23-27) la *visión estática*, aquella que parece influir en la mayoría de la gente común y en los estudiantes, consiste en que la ciencia es una actividad que aporta al mundo información sistematizada. El trabajo del científico es descubrir nuevos hechos y agregarlos al cuerpo ya existente de información. Se concibe incluso a la ciencia como un conjunto de hechos. Desde esta perspectiva la ciencia es también una forma de explicar los fenómenos observados. El énfasis está entonces en el estado *actual del conocimiento* y en *la adición* que se le hace, así como en el conjunto de leyes, teorías, hipótesis y principios actuales.

La *visión dinámica*, por otro lado, considera a la ciencia más como una actividad que como aquello que realizan los científicos. El estado actual del conocimiento es importante, por supuesto, pero lo es en tanto que constituye la base para futuras teorías e investigaciones científicas. A esto se le ha llamado *visión heurística*. La palabra *heurística* significa "que sirve para descubrir o revelar", y ahora tiene la connotación de autodescubrimiento. Un método heurístico de enseñanza, por ejemplo, subraya la importancia de que los estudiantes descubran las cosas por sí mismos. La visión heurística en la ciencia se centra en la teoría y esquemas conceptuales interconectados que resultan fructíferos para investigaciones posteriores. Un énfasis heurístico implica un énfasis en el descubrimiento.

Es esta visión heurística de la ciencia lo que la distingue en buena medida de la ingeniería y la tecnología. Con base en esta corazonada heurística el científico da un salto riesgoso. Como dice Polanyi (1958/1974, p. 123), "Es el impulso por el cual ganamos pie en la otra orilla de la realidad. En tales casos el científico tiene que apostar, pieza a pieza, toda su vida profesional". Michel (1991, p. 23) agrega: "quien teme ser malinterpretado, y por esta razón estudia un método científico 'seguro' o 'cierto', no debe involucrarse en investigación científica alguna". La visión heurística también puede llamarse solución de problemas, pero el énfasis está en lo imaginativo y no en la solución rutinaria de problemas. La visión heurística en la ciencia enfatiza la resolución de problemas, más allá de los hechos y conjuntos de información. Éstos resultan importantes para el científico heurístico porque le ayudan a encaminarse hacia teorías, descubrimientos e investigaciones futuras.

Al evitar aún una definición directa de la ciencia —pero ciertamente implicándola— ahora se abordará la función de la ciencia. Aquí se tienen dos visiones distintas. La persona práctica, generalmente quien no es científico, considera a la ciencia como una disciplina o actividad encaminada a mejorar las cosas, a generar progresos. También algunos científi-

cos asumen esta postura. La función de la ciencia, desde esta perspectiva, consiste en hacer descubrimientos, conocer hechos y avanzar el conocimiento con el fin de mejorar las cosas. Las ramas de la ciencia que claramente pertenecen a este género reciben un apoyo amplio y fuerte, como el caso de la investigación médica y meteorológica. El criterio de utilidad práctica y "resultado" son relevantes en esta perspectiva, en especial en la investigación educativa (véase Kerlinger, 1977; Bruno, 1972).

Un punto de vista muy diferente sobre la función de la ciencia está bien expresado por Braithwaite (1953/1996, p. 1):

> La función de la ciencia... consiste en establecer leyes generales sobre el comportamiento de eventos empíricos u objetos en los que la ciencia en cuestión está interesada, para así permitirnos conectar nuestro conocimiento de eventos conocidos por separado y hacer predicciones confiables de eventos aún desconocidos.

La conexión entre esta perspectiva de la función de la ciencia y la dinámica-heurística antes discutida es obvia, excepto que se ha agregado un elemento muy importante: el establecimiento de leyes generales —o teoría si se quiere—. Si hemos de comprender la investigación conductual moderna junto con sus fortalezas y debilidades debemos explorar los elementos de la declaración de Braithwaite. Lo hacemos al considerar el fin de la ciencia, la explicación científica y el papel e importancia de la teoría.

Sampson (1991) analiza dos puntos de vista opuestos de la ciencia. Existe una perspectiva convencional o tradicional y una perspectiva sociohistórica. La convencional percibe a la ciencia como un espejo de la naturaleza o como una vitrina de cristal transparente que presenta la naturaleza sin sesgo ni distorsión. El objetivo en este caso es describir con el máximo grado de exactitud cómo es el mundo en realidad. Sampson establece que la ciencia constituye un árbitro objetivo. Su trabajo es "resolver los desacuerdos y distinguir qué es cierto y correcto, de aquello que no lo es". Cuando la visión convencional de la ciencia es incapaz de resolver la disputa, esto sólo significa que hay datos o información insuficientes para hacerlo. Los convencionalistas, sin embargo, consideran que sólo es cuestión de tiempo para que la verdad salte a la vista.

La visión sociohistórica concibe a la ciencia como una historia. Los científicos son narradores. La idea es que la realidad sólo puede ser descubierta a través de las historias que se cuentan acerca de ella. Este enfoque es diferente de la visión tradicional-convencional en tanto que no hay un árbitro neutral. Cada historia está sazonada por la orientación del narrador. Como resultado, no hay una historia verdadera única. La interpretación del autor sobre la presentación de Sampson que compara estos dos aspectos se muestra en la tabla 1.1.

Aun cuando Sampson proporciona estas dos visiones de la ciencia a la luz de la psicología social, su presentación es aplicable en todas las áreas de las ciencias del comportamiento.

Los objetivos de la ciencia, explicación científica y teoría

El objetivo básico de la ciencia es la teoría. Quizás, dicho de forma menos críptica, el fin básico de la ciencia es explicar los fenómenos naturales. Tales explicaciones se llaman "teorías". En lugar de tratar de explicar cada una de las conductas de los niños por separado, el psicólogo científico busca explicaciones generales que abarquen y conjunten muchas conductas diferentes. En vez de intentar explicar los métodos que usan los niños para resolver los problemas aritméticos, por ejemplo, el científico busca explicaciones genera-

TABLA 1.1 *Los dos puntos de vista de Sampson acerca de la ciencia y la psicología social*

	Tradicional (Cuantatativo)	No tradicional (Sociohistórico) (Cuantatativo)
Objetivo primario	Describir la realidad de las interacciones y funciones humanas y sociales.	Describir la variedad de experiencias y actividades humanas y sociales a través de la información histórica y social y de los papeles que desempeñan en la vida humana.
Posición filosófica	La realidad puede ser descubierta de forma independiente por observadores neutros. La realidad puede ser apreciada sin ocupar ninguna posición particular que genere sesgos.	La realidad puede ser descubierta sólo desde algún punto de vista: el observador, entonces, siempre está posicionado.
Enunciado metafórico	Se puede percibir a la ciencia como un espejo. Está diseñada para reflejar las cosas tal como son en realidad.	La ciencia se percibe como un contador de historias que proporciona versiones diferentes o personales de la realidad.
Consideraciones metodológicas	Los métodos se crean y utilizan para controlar o eliminar factores que debilitarían la habilidad del investigador para descubrir la verdadera forma de la realidad.	Amplios factores históricos y sociales moldean la comprensión que el investigador tiene de la realidad. Los métodos pueden generar un entendimiento más rico y profundo de la realidad con base en el encuentro de diferentes versiones que la gente utiliza para comprender sus vidas.

les de todos los tipos de solución de problemas. Esto podría llamarse una teoría general de solución de problemas.

Este análisis sobre el objetivo básico de la ciencia como teoría puede resultar extraño al estudiante al que quizá se le ha inculcado la idea de que las actividades humanas han de producir resultados prácticos. Si dijéramos que el objetivo de la ciencia es el progreso de la humanidad, la mayoría leería las palabras con rapidez y las aceptaría. Pero el objetivo *básico* de la ciencia *no* es el progreso de la humanidad: es la teoría. Por desgracia, este enunciado vasto y realmente complejo no es fácil de entender. Aún así, debemos tratar de comprenderlo porque es importante. Hay una explicación más amplia de este punto en el capítulo 16 de Kerlinger (1979).

Otros objetivos de la ciencia que se han mencionado son: la explicación, comprensión, predicción y el control. Sin embargo, si aceptamos la teoría como el fin supremo de la ciencia, la explicación y el entendimiento se convierten en subobjetivos del objetivo fundamental debido a la definición y naturaleza de la teoría: ***una teoría es un conjunto de constructos (conceptos) interrelacionados, definiciones y proposiciones que presentan una visión sistemática de los fenómenos al especificar las relaciones entre variables con el propósito de explicar y predecir los fenómenos.***

Esta definición indica tres cosas: 1) una teoría es un conjunto de proposiciones constituidas por constructos definidos e interrelacionados, 2) una teoría establece las interre-

laciones entre un conjunto de variables (constructos) y, al hacerlo, presenta una visión sistemática del fenómeno descrito por las variables, y 3) una teoría explica fenómenos al especificar qué variables están relacionadas con cuáles otras y de qué forma están relacionadas. De esta manera permiten al investigador hacer predicciones de ciertas variables a partir de otras. Uno podría, por ejemplo, contar con una teoría sobre el fracaso escolar. Nuestras variables podrían ser inteligencia, aptitudes numérica y verbal, ansiedad, clase social, estado nutricional y motivación de logro. El fenómeno a ser explicado es, por supuesto, el fracaso escolar, o más precisamente, el rendimiento escolar. El fracaso escolar puede concebirse como un extremo del continuo de rendimiento escolar, mientras en el otro estaría el éxito escolar. El fracaso escolar se explica a partir de las relaciones especificadas entre cada una de las siete variables y el fracaso escolar, o por la combinación de las siete variables y el fracaso escolar. El científico, al utilizar con éxito este conjunto de constructos puede "entender" el fracaso escolar, es capaz de "explicarlo" y, al menos en alguna medida, "predecirlo".

Resulta evidente que la explicación y predicción pueden ser incluidas en una teoría. La misma naturaleza de una teoría consiste en la explicación de los fenómenos observados. Tomemos, por ejemplo, la teoría del reforzamiento. Una proposición sencilla que se deriva de ella es: si una respuesta es premiada (reforzada) cuando ocurre, tenderá a repetirse. El primer psicólogo científico que formuló esta proposición lo hizo como una forma de explicar la ocurrencia repetida de respuestas que había observado. ¿Por qué ocurrieron y volvieron a presentarse con una regularidad confiable? Porque fueron recompensadas. Aunque esto constituye una explicación puede no resultar satisfactoria para mucha gente. Algunos pueden preguntar por qué el premio aumenta la probabilidad de que ocurra una respuesta. Una teoría completa contendría la explicación. Sin embargo, hoy en día no existe una respuesta realmente satisfactoria. Todo lo que podemos decir es que con un alto grado de probabilidad, el reforzamiento de una respuesta hace que sea más probable que ocurra una y otra vez. (véase Nisbett y Ross, 1980). En otras palabras, las proposiciones de una teoría, las declaraciones de relaciones, constituyen la explicación, en cuanto a la teoría de fenómenos naturales observados.

En cuanto a la predicción y el control puede decirse que los científicos no tienen que estar realmente involucrados en la explicación y la comprensión. Sólo la predicción y el control son necesarios. Quienes proponen esta postura dirían que el poder predictivo marca qué tan adecuada resulta. Si al utilizar una teoría somos capaces de predecir con éxito, entonces la teoría se confirma y eso es suficiente, no es necesario buscar más explicaciones subyacentes. En tanto podemos predecir con confiabilidad podemos controlar, ya que el control se deriva de la predicción.

La perspectiva de la predicción en la ciencia tiene validez. Pero por lo que a este libro compete, la predicción se considera un aspecto de la teoría. Por su propia naturaleza, una teoría predice; cuando de las proposiciones originales de una teoría deducimos otras más complejas, en esencia estamos "prediciendo". Cuando explicamos fenómenos observados siempre establecemos una relación entre, por ejemplo, la clase *A* y la clase *B*. La explicación científica reside en especificar las relaciones entre una clase de eventos empíricos y otra, bajo ciertas circunstancias. Decimos: si *A*, entonces *B*; en donde *A* y *B* se refieren a clases de objetos o eventos.[2] Pero esto constituye una predicción, la predicción de *A* acer-

[2] Enunciados de la forma: "si *p*, entonces *q*" se llaman *enunciados condicionales* en lógica y son la base del cuestionamiento científico. Ellos y los conceptos o variables que incluyen son el ingrediente central de las teorías. El fundamento lógico del cuestionamiento científico que subyace a gran parte del razonamiento de este libro está resumido en Kerlinger (1977).

ca de *B*. Así, una explicación teórica implica una predicción, lo que nos lleva de nuevo a la idea de que la teoría es el objetivo final de la ciencia. Todo lo demás se deriva de la teoría.

Nuestra intención no es desacreditar o denigrar la investigación que no está específica y conscientemente orientada a la teoría. Muchas investigaciones valiosas en ciencias sociales y educativas se interesan por el objetivo más constreñido de encontrar relaciones específicas; es decir, el solo hecho de descubrir una relación forma parte de la ciencia. Pero las relaciones más útiles y satisfactorias son, en última instancia, aquéllas que tienen el máximo grado de generalización, las que están ligadas a otras relaciones en una teoría.

El concepto de generalidad es importante. En tanto que son generales las teorías se aplican a una variedad de fenómenos y a mucha gente de diversos lugares. Una relación específica, por supuesto, tiene un espectro de aplicación más reducido. Si por ejemplo, uno encuentra que la ansiedad al responder pruebas está relacionada con el desempeño, por interesante e importante que sea este hallazgo, tiene menor aplicabilidad y es menos comprendido que el descubrimiento de una relación en una red de variables interrelacionadas que forman parte de una teoría. Por lo tanto, los objetivos de investigación modestos, limitados y específicos son buenos, pero los de la investigación teórica son mejores, entre otras razones porque son más generales y aplicables a un amplio margen de situaciones. Además, cuando existe tanto una teoría simple como una compleja, y ambas dan cuenta de los hechos de forma igualmente efectiva, se prefiere la explicación sencilla (Navaja de Occam).* De aquí que en la discusión sobre la posibilidad de generalizar, una buena teoría también es parsimoniosa. Sin embargo, una cantidad de teorías incorrectas sobre la enfermedad mental persiste a causa de este atributo: algunos aún creen que los individuos están poseídos por demonios. Tal explicación es simple si se compara con las médicas y/o psicológicas.

Las teorías son explicaciones tentativas. Se evalúa cada teoría empíricamente para determinar qué tan bien predice los nuevos hallazgos. Las teorías pueden usarse para guiar un plan de investigación al generar hipótesis susceptibles de ser probadas, y para organizar hechos obtenidos al probar estas hipótesis. Una buena teoría es aquella que no se ajusta a todas las observaciones. Uno debería ser capaz de encontrar una ocurrencia que la contradiga. La teoría de Blondlot de los rayos N es un ejemplo de una teoría pobre. Blondlot expresó que toda la materia emitía rayos N (Weber, 1973). Aunque se demostró más tarde que los rayos N no existían, Barber (1976) indicó que casi 100 artículos se publicaron sobre estos rayos en un año, en Francia. Blondlot incluso desarrolló equipo complicado para observar rayos N. Los científicos que declararon haber observado rayos N sólo fortalecieron la teoría y hallazgos de Blondlot. Sin embargo, cuando alguien dijo no haber visto los rayos N, Blondlot declaró que sus ojos no eran lo suficientemente sensibles, o que no había ajustado el instrumento de forma apropiada. Ningún resultado se consideró evidencia en contra de la teoría. En tiempos más recientes otra teoría falsa que tomó más de 75 años derrocar fue la relativa al origen de la úlcera péptica. En 1910 Schwartz (citado en Blaser, 1996) declaró haber establecido firmemente la causa de las úlceras: los ácidos gástricos. En años posteriores investigadores médicos dedicaron su tiempo y energía al tratamiento de las úlceras a través del desarrollo de medicamentos para neutralizar o bloquear los ácidos, los cuales nunca tuvieron mucho éxito y resultaban costosos. Sin embargo, en 1985 J. Robin Warren y Barry Marshall (citados en Blaser, 1996) descubrieron que el *helioc bacter pylori* era la causa real de las úlceras gástricas. Casi todos los casos de este

* N. del T. La *navaja de Occam* (Ockamm's razor) procede de William de Occam (1285-1349), filósofo inglés que formuló la máxima: *entia non sunt multiplicanda praeter necessitatem*, esto es: *las suposiciones que explican un fenómeno no deberán ser multiplicadas más allá de lo necesario*. Sir William Hamilton en 1853 la denominó *ley de la parsimonia*.

tipo de úlcera fueron tratados con éxito por medio de antibióticos y a un precio considerablemente más bajo. Durante 75 años ningún resultado se tomó como evidencia contra la teoría del ácido-estrés de las úlceras.

La investigación científica: definición

Es más fácil definir la investigación científica que la ciencia. Sin embargo, no sería sencillo lograr un acuerdo de científicos e investigadores en relación con una definición. Aun así intentaremos una: *la investigación científica es una investigación sistemática, controlada, empírica, amoral, pública y crítica de fenómenos naturales. Se guía por la teoría y las hipótesis sobre las presuntas relaciones entre esos fenómenos.* Esta definición requiere poca explicación en tanto que es una declaración condensada y formalizada de muchos aspectos que ya se han presentado o que abordaremos pronto. Sin embargo, es necesario enfatizar dos puntos.

Primero, cuando decimos que la investigación científica es sistemática y controlada queremos decir, de hecho, que la investigación es tan ordenada que los investigadores pueden tener una confianza crítica en los resultados. Como se verá más adelante, las observaciones de la investigación científica son estrictamente disciplinadas. Más aún, entre las muchas explicaciones alternativas de un fenómeno, todas menos una se rechazan de forma sistemática. Así, uno puede tener mayor confianza en que una relación sometida a prueba es tal como es, que si no se hubieran controlado las observaciones y desechado las posibilidades alternativas. En algunos casos es posible establecer una relación de causa-efecto.

Segundo, la investigación científica es empírica. Si el científico cree que algo se da de cierta forma debe demostrarlo de un modo u otro por medio de una prueba independiente externa. En otras palabras, las consideraciones subjetivas deben ser verificadas contra una realidad objetiva. Los científicos siempre deben presentar sus nociones ante el tribunal del cuestionamiento y prueba empíricos. Los científicos son sumamente críticos de sus propios resultados y de los de las investigaciones de los demás. Cada científico que redacta un informe de investigación cuenta con otros científicos que leen lo que él escribe a lo largo de todo el proceso. Aunque es fácil errar, exagerar, generalizar de más al redactar uno mismo su propio trabajo, no es fácil escapar al escrutinio de otros científicos que vigilan la tarea.

En ciencia existe la revisión de pares. Esto significa que otros con igual capacidad y conocimiento son llamados para evaluar el trabajo del científico antes de que sea publicado en revistas especializadas. En este punto hay tantos aspectos positivos como negativos. Es a través de esta revisión de pares que se han descubierto estudios fraudulentos. El ensayo escrito por R.W. Wood (1973) acerca de sus experiencias con el profesor francés Blondlot, sobre la inexistencia de los rayos *N*, aporta una clara demostración de estas revisiones, que resultan buenas para la ciencia y promueven la calidad en la investigación. El sistema, sin embargo, no es perfecto. Hay ocasiones en que la evaluación de pares se ha volcado contra la ciencia. Esto está documentado a través de la historia con personas como Kepler, Galileo, Copérnico, Jenner y Semelweiss. Las ideas de estos individuos no fueron populares entre colegas. Más recientemente, en psicología, el trabajo de John García sobre las restricciones biológicas del aprendizaje fue contrario a los de sus colegas. García consiguió publicar sus hallazgos en una revista (*Bulletin of the Psychonomic Society*) que no tenía evaluación de pares. Algunos investigadores que leyeron y replicaron su trabajo lo encontraron valioso. En la gran mayoría de los casos la revisión de colegas resulta benéfica para la ciencia.

Tercero, el conocimiento obtenido científicamente no está sujeto a una evaluación moral. Los resultados no se juzgan por "malos" ni "buenos", sino en términos de validez y confiabilidad. Sin embargo, el método científico está sujeto a principios de moralidad; es decir, que el científico es responsable de los métodos utilizados para obtener el conocimiento científico. En psicología, los códigos de ética se establecen para proteger a quienes están bajo estudio. La ciencia es una aventura compartida. La formación científica está disponible para todos y el método científico es bien conocido y está a la mano de quienes eligen usarlo.

El enfoque científico

El enfoque científico es una forma especial y sistematizada del pensamiento y del cuestionamiento reflexivos. Dewey (1933/1991), en su influyente trabajo *How We Think (Cómo Pensamos)*, delineó un paradigma general del cuestionamiento. La presente discusión del enfoque científico se basa en gran parte en el análisis de Dewey.

Problema-obstáculo-idea

El científico puede experimentar dificultades para entender, una vaga inquietud acerca de los fenómenos observados y no observados, una curiosidad sobre por qué algo es de la forma en que se presenta. El primer paso —y el más importante— es tener que sacar a la luz una idea, expresar el problema de alguna forma razonablemente manejable. Nunca o rara vez el problema surgirá por completo en esta etapa. El científico debe esforzarse con él, luchar con él y vivir con él. Dewey (1933/1991, p. 108) dice: "Hay una situación problemática, perpleja, irritante, donde la dificultad se encuentra a todo lo largo y ancho de ella, afectándola como un todo." Más tarde o más temprano, explícita o implícitamente, el científico definirá el problema, incluso si su expresión resulta incipiente y tentativa. El científico intelectualiza, como Dewey (p. 109) señala "aquello que al principio es meramente una cualidad *emocional* de toda la situación" (las itálicas son agregadas). En algunos aspectos ésta es la parte más difícil e importante de todo el proceso. Sin algún tipo de definición del problema, el científico difícilmente podrá seguir adelante y esperar que su trabajo sea fructífero. Para algunos investigadores la idea puede provenir de la conversación con un colega, o de la observación de un fenómeno curioso. La idea es que el problema por lo general se inicia con un pensamiento vago o no científico, o con presentimientos no sistemáticos. Después seguirán pasos más refinados.

Hipótesis

Después de intelectualizar el problema, de referirse a experiencias pasadas para posibles soluciones, de observar fenómenos relevantes, el científico puede formular una hipótesis. Una hipótesis es una declaración conjetural, una proposición tentativa acerca de la relación entre dos o más fenómenos o variables. Nuestro científico dirá, "si ocurre tal y tal, entonces resultará tal y tal".

Razonamiento-deducción

Este paso o actividad con frecuencia pasa inadvertido o es poco enfatizado. Quizás es la parte más importante del análisis de Dewey sobre el pensamiento reflexivo. El científico

deduce las consecuencias de la hipótesis que él mismo ha formulado. Conant (1951), al hablar acerca del surgimiento de la ciencia moderna, indica que el nuevo elemento que se aportó en el siglo XVII fue el uso del razonamiento deductivo. Aquí es donde la experiencia, el conocimiento y la perspicacia son importantes.

Con frecuencia el científico, al deducir las consecuencias de una hipótesis formulada, llegará a un problema muy diferente del original. Por otro lado, las deducciones pueden hacer creer que el problema no puede resolverse con las herramientas técnicas actuales. Por ejemplo, antes que la estadística moderna se desarrollara, algunos problemas de investigación del comportamiento eran irresolubles. Era difícil, si no es que imposible, probar dos o tres hipótesis interdependientes de forma simultánea, y también era casi imposible probar el efecto interactivo de variables. Ahora hay razón para pensar que ciertos problemas no pueden resolverse a menos que se les aborde de forma multivariada. Un ejemplo de esto es la relación entre los métodos de enseñanza con el aprovechamiento escolar y otras variables. Es probable que los métodos de enseñanza, *per se*, no difieran mucho si sólo se estudian sus efectos simples. Los métodos de enseñanza trabajan en forma diferente bajo distintas condiciones, con diversos maestros y con alumnos variados. Se dice que los métodos "interactúan" con las condiciones y características de los docentes y de los estudiantes. Simon (1987) presentó otro ejemplo: un estudio de investigación sobre entrenamiento de pilotos propuesto por Williams y Adelson en 1954 no se podía realizar usando los métodos tradicionales de experimentación. El estudio proponía examinar 34 variables y su influencia en el entrenamiento de los pilotos. Con el uso de métodos tradicionales de investigación el número de variables bajo estudio era abrumador. Alrededor de 20 años después, Simon (1976) y Simon y Roscoe (1984) demostraron la forma en que se podía abordar con efectividad tales estudios usando diseños económicos de megafactor. Un ejemplo puede ayudarnos a entender este paso de razonamiento-deducción.

Suponga que un investigador está intrigado por la conducta agresiva. El investigador se pregunta por qué las personas son con frecuencia agresivas en situaciones donde este comportamiento puede ser inapropiado. La observación personal nos lleva a suponer que la conducta agresiva parece ocurrir cuando las personas han experimentado dificultades de uno u otro tipo. (Nótese la vaguedad del problema en este punto). Después de pensar por un tiempo, revisar la literatura para obtener algunas claves y llevar a cabo más observaciones, se formula la hipótesis: la frustración conduce a la agresividad. *La frustración* se define como el obstáculo para alcanzar una meta y la *agresividad* como la conducta caracterizada por ataque físico o verbal a otras personas u objetos.

Lo que sigue es una declaración como ésta: si la frustración conduce a la agresividad, entonces deberíamos encontrar un alto grado de agresividad entre los niños de escuelas restrictivas, que no les permiten mucha libertad ni posibilidad de expresión. De forma similar, en situaciones sociales difíciles, suponiendo que son frustrantes, esperaríamos más agresividad de lo "común". Si seguimos el razonamiento, si les diéramos a sujetos experimentales problemas interesantes para resolver, y después evitáramos que los solucionaran, podemos predecir algún tipo de conducta agresiva. En pocas palabras, el proceso de trasladarnos de un contexto amplio a una situación más específica se llama *razonamiento deductivo*.

El razonamiento puede, como se indicó antes, cambiar el problema. Podemos comprender que el problema inicial era sólo un caso especial de un problema más amplio, fundamental e importante. Podemos por ejemplo, iniciar con una hipótesis más limitada: las situaciones escolares restrictivas conducen a negativismo en los niños. Después podemos generalizar el problema a: la frustración induce la agresividad. Aun cuando es una forma diferente de pensamiento de lo arriba discutido, es importante, por lo que casi podríamos llamar su calidad heurística. El razonamiento puede ayudar a dirigirnos hacia

problemas más amplios, más básicos y por tanto más significativos, así como a aportar implicaciones operacionales (susceptibles de ser probadas) de la hipótesis original. Este tipo se llama *razonamiento inductivo*. Parte de un hecho particular hacia un enunciado general o hipótesis. Si uno no es cuidadoso este método puede inducir un razonamiento deficiente debido a su tendencia natural de excluir datos que no se ajustan a la hipótesis. El método de razonamiento inductivo tiende a buscar más datos de apoyo que a refutar evidencias.

Considere el estudio clásico de Peter Wason (Wason y Johnson-Laird, 1972) que ha suscitado un gran interés (Hoch 1986; Klayman y Ha, 1987). En este estudio se le pidió a estudiantes descubrir la regla que el experimentador tenía en mente al generar una secuencia de números. Un ejemplo fue generar una regla de la siguiente serie: "3, 5, 7". Se les dijo a los alumnos que podrían preguntar acerca de otras secuencias y que recibirían retroalimentación en cada serie sobre si se ajustaron o no a la regla pensada por el experimentador. Cuando los estudiantes se sintieran seguros podrían externar la regla. Algunos estudiantes indicaron: "9, 11, 13", y les dijeron que esta secuencia se ajustaba a la regla. Después siguieron con "15, 17, 19" y otra vez les respondieron que la serie correspondía. Los estudiantes entonces presentaron su respuesta: "la regla es tres números nones consecutivos", pero se les dijo que ésta *no* era la regla. Después de varias secuencias más propusieron contestaciones tales como: "números con incrementos de dos en dos", o bien "números nones con incrementos de valor dos". En cada uno de los casos se les indicó que ésa no era la regla que el experimentador había pensado. La regla que el experimentador tenía en mente era: "tres números positivos crecientes cualesquiera". Si los estudiantes hubieran propuesto las secuencias "8, 9, 10" o "1, 15, 4500" les habrían dicho que estos números también se ajustaban a la regla. Donde los estudiantes cometieron el error fue en probar sólo los casos que se ajustaban a la primera secuencia propuesta y que confirmaba su hipótesis.

Aunque simplificado en exceso, el estudio de Wason demostró lo que puede pasar en la investigación científica real. Un científico puede fácilmente condenarse a repetir el mismo tipo de experimento que siempre apoye la hipótesis.

Observación-prueba-experimento

A estas alturas debe quedar claro que la fase de observación-prueba-experimento forma sólo parte de todo el proceso científico. Si el problema ha sido bien planteado, la o las hipótesis se han formulado de manera adecuada y las implicaciones de las hipótesis se han deducido con cuidado, se puede presumir en este paso que el investigador es competente desde el punto de vista técnico.

La esencia de la prueba de hipótesis consiste en probar la *relación* expresada por la hipótesis. No se prueban las variables como tales, sino la relación entre ellas. La observación, la prueba y la experimentación tienen un propósito fundamental: probar empíricamente la relación del problema. Probar sin saber —al menos en alguna medida— *qué* y *por qué* uno evalúa implica un disparate. El hecho de sólo plantear un problema vago, por ejemplo, "¿cómo afecta al aprendizaje la educación abierta?" para después evaluar alumnos en las escuelas que dicen ser diferentes en cuanto a su "nivel de apertura", o preguntarnos: "¿cuáles son los efectos de la disonancia cognitiva?" para luego crear disonancia a través de manipulaciones experimentales y buscar los supuestos efectos, todo esto sólo podría llevarnos a información cuestionable. De forma similar, decir que se estudiarán los procesos de atribución sin realmente saber lo que se hace, o sin establecer relaciones entre las variables, constituye investigación sin sentido.

Otro aspecto importante es que por lo general no se prueba la hipótesis de manera directa. Como se indicó en el paso previo sobre el razonamiento, probamos las implicaciones que deducimos de las hipótesis. Nuestra hipótesis a probar puede ser: "los sujetos a quienes se les instruye para que eviten pensamientos indeseados estarán más preocupados con ellos que aquellos a los que se da una distracción". Ésta se dedujo de una hipótesis más amplia y general: "Cuantos mayores sean los esfuerzos para suprimir una idea, mayor será la preocupación sobre esta idea." No probamos "la supresión de ideas" o "la preocupación" sino la relación entre ellas, en este caso, la relación entre supresión de pensamientos indeseados y el nivel de preocupación (véase Wegner, Schneider, Carter y White, 1987; Wegner, 1989).

Dewey enfatizó que la secuencia temporal del pensamiento o cuestionamiento reflexivos no está fija. Podemos repetir y enfatizar lo que dijo en nuestro propio marco: los pasos del enfoque científico no están ordenados de manera impecable. El primer paso no se completa perfectamente antes de iniciar el segundo. Más aún, podemos hacer pruebas antes de deducir de forma adecuada las implicaciones de la hipótesis. La hipótesis, en sí misma, puede necesitar una mayor elaboración o refinamiento como resultado de las implicaciones deducidas de ella. Con frecuencia veremos que las hipótesis y su expresión parecen inadecuadas una vez que se deducen implicaciones a partir de ellas. Cuando una hipótesis es muy vaga se presenta la dificultad común de observar que una deducción es tan buena como otra; esto es, la hipótesis puede no conducir a pruebas precisas.

Retroalimentar el problema, la hipótesis y finalmente, la teoría de los resultados de la investigación es de la mayor importancia. Los teóricos e investigadores del aprendizaje, por ejemplo, con frecuencia han modificado sus teorías e investigaciones como resultado de hallazgos experimentales (véase Malone, 1991; Schunk, 1996; Hergenhahn, 1996). Teóricos e investigadores han estudiado los efectos del entorno y del entrenamiento tempranos en el desarrollo posterior. Kagan y Zentner (1996) revisaron los resultados de 70 estudios sobre las relaciones entre las experiencias de la edad temprana y la psicopatología en la vida adulta. Encontraron que es posible predecir la delincuencia juvenil a partir de la cantidad de impulsividad detectada en la etapa preescolar. Lynch, Short y Chua (1995) hallaron que el procesamiento musical estaba influido por la estimulación perceptual experimentada por el niño entre los seis meses y el año. Éstas y otras investigaciones han generado evidencia variada que converge en este problema de extrema importancia teórica y práctica. Una parte esencial de la investigación científica es el esfuerzo constante por replicar y verificar los hallazgos, por corregir la teoría con base en la evidencia empírica y por encontrar mejores explicaciones para los fenómenos naturales. Se puede incluso aseverar que la ciencia tiene un aspecto cíclico. Un investigador encuentra, por ejemplo, que *A* está relacionada con *B* de tal o cual forma. Entonces se conduce más investigación para determinar bajo qué otras condiciones *A* está relacionada de forma similar a *B*. Otros investigadores desafían esta teoría e investigación y ofrecen sus propias evidencias y explicaciones. El investigador original, se espera, modificará su trabajo a la luz de los nuevos datos: el proceso nunca termina.

Resumamos el llamado enfoque científico de la investigación. En primera instancia existe una duda, una barrera, una situación indeterminada que debe determinarse. El científico experimenta dudas vagas, disturbios emocionales e ideas incipientes. Hay un esfuerzo por formular el problema aunque sea de forma inadecuada. El científico entonces revisa la literatura y busca en su propia experiencia y en la de otros. Es frecuente que el investigador tenga que esperar un momento de inventiva: puede ser que ocurra, puede ser que no. Una vez que se cuenta con el problema formulado, con la o las preguntas básicas expresadas de manera apropiada, el resto es más fácil. Entonces, la hipótesis se construye, después de lo cual se deducen las implicaciones empíricas. Durante este proceso el proble-

ma original, y por supuesto la hipótesis original, pueden cambiarse, hacerse más generales o reducirse. Incluso pueden abandonarse. Más tarde se prueba la relación expresada por la hipótesis por medio de la observación y la experimentación. Con base en la evidencia procedente de la investigación, la hipótesis se apoya o rechaza. Esta información después retroalimenta al problema original, que se conserva o modifica de acuerdo con la evidencia. Dewey señaló que una fase del proceso puede expandirse y ser de gran importancia, otra puede reducirse, y puede haber más o menos pasos involucrados. La investigación rara vez es un asunto ordenado. De hecho, resulta mucho más desordenado que lo que la discusión anterior pueda sugerir. El orden y el desorden, sin embargo, no son de primera importancia. Lo que *sí* es importante es la racionalidad controlada de la investigación científica como un proceso de indagación reflexiva, la naturaleza interdependiente de las partes del proceso y la suprema importancia del problema y su enunciado.

RESUMEN DEL CAPÍTULO

1. Para comprender la compleja conducta humana es necesario entender el lenguaje y enfoque científicos.
2. La ciencia es una extensión sistemática y controlada del sentido común. Hay cinco diferencias entre el sentido común y la ciencia:
   a) La ciencia utiliza esquemas conceptuales y estructuras teóricas.
   b) La ciencia prueba de forma sistemática y empírica las teorías e hipótesis.
   c) La ciencia intenta controlar posibles causas extrañas.
   d) La ciencia busca relaciones de manera consciente y sistemática.
   e) La ciencia excluye explicaciones metafísicas (no demostrables).
3. Los cuatro métodos del conocimiento de Peirce son:
   a) Método de la tenacidad —influenciado por las creencias pasadas ya establecidas—.
   b) Método de autoridad —determinado por el peso de la tradición o la sanción pública—.
   c) Método *a priori* (también conocido como método de la intuición) —una natural inclinación hacia la verdad—.
   d) Método de la ciencia —autocorrectivo; los conceptos son objetivos y susceptibles de ser probados—.
4. Los estereotipos de la ciencia han limitado la comprensión de esta actividad por parte del público.
5. Visión y función de la ciencia
   a) Una visión estática considera a la ciencia como proveedora de información para el mundo; la ciencia aporta al cuerpo de información y al estado actual del conocimiento.
   b) La visión dinámica está interesada en la actividad de la ciencia (lo que hacen los científicos). Con ello aparece la visión heurística de la ciencia, que tiene un carácter de autodescubrimiento. La ciencia asume riesgos y resuelve problemas.
6. Los objetivos de la ciencia son:
   a) Generar teoría y explicar los fenómenos naturales.
   b) Promover la comprensión y desarrollar predicciones.
7. Una teoría tiene tres características:
   a) Posee un conjunto de propiedades con base en constructos definidos e interrelacionados.

b) Determina de forma sistemática las interrelaciones entre un grupo de variables.
c) Explica fenómenos.

8. La investigación científica es una búsqueda sistemática, controlada, empírica y crítica de fenómenos naturales. La teoría e hipótesis acerca de relaciones que se presume existen entre tales fenómenos la orienta. Es pública y amoral.
9. El enfoque científico, de acuerdo con Dewey, está integrado por:
   a) Problema-obstáculo-idea —formular el problema o pregunta de investigación a resolver—.
   b) Hipótesis —formular un enunciado conjetural acerca de la relación entre fenómenos o variables—.
   c) Razonamiento-deducción —el científico deduce las consecuencias de la hipótesis, lo cual puede conducir a un problema más significativo y generar ideas sobre cómo las hipótesis pueden ser probadas en términos observables—.
   d) Observación-prueba-experimento —constituyen la fase de recolección y análisis de datos—. Los resultados de la investigación se relacionan una vez más con el problema.

SUGERENCIAS DE ESTUDIO

Una parte de este capítulo se presta a gran controversia. Algunos pensadores aceptan esos puntos de vista, mientras otros los rechazan. La revisión de literatura selecta contribuye a mejorar la comprensión de la ciencia y de su propósito, de la relación entre ciencia y tecnología, y las diferencias entre investigación básica y aplicada. Esas lecturas pueden constituir la base de las discusiones en clase. En el libro del primer autor, *Behavioral Research: A Conceptual Approach* (Nueva York: Holt, Rinehart y Winston, 1979, capítulos 1, 15 y 16) es posible encontrar una amplia presentación de aspectos controvertidos de la ciencia, en especial, de la ciencia del comportamiento. Se ha publicado una gran cantidad de artículos de buena calidad en ciencia e investigación, en revistas científicas y en libros de filosofía de la ciencia. Aquí presentamos algunos. También se incluye un informe especial en *Scientific American*. Todos son relevantes para el contenido de este capítulo.

Barinaga, M. (1993). Philosophy of science: Feminists find gender everywhere in science. *Science, 260*, 392-393. Analiza la dificultad de separar puntos de vista culturales de las mujeres y la ciencia. Señala a la ciencia como un campo predominantemente masculino.

Hausheer, J., Harris, J. (1994). In search of a brief definition of science. *The Physics Teacher, 32*(*5*), 318. Menciona que cualquier definición de ciencia debe incluir guías para evaluar teorías e hipótesis como científicas o no científicas.

Holton, G. (1996). The controversy over the end of science. *Scientific American, 273*(*10*), 191. Este artículo versa sobre dos campos del pensamiento: los linealistas y los cíclicos. Los linealistas tienen una perspectiva más convencional de la ciencia; los cíclicos conciben a la ciencia como un proceso que se degenera de forma interna.

Horgan, J. (1994). Anti-omniscience: An eclectic gang of thinkers pushes at knowledge's limits. *Scientific American, 271*, 20-22. Analiza los límites de la ciencia.

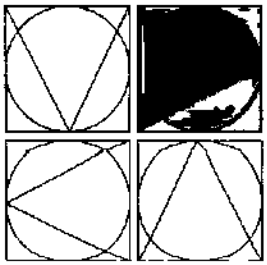
Horgan, J. (1997). *The end of science*. Nueva York: Broadway Books.

Miller, J.A. (1994). Postmodern attitude toward science. *Bioscience, 41*(*6*), 395. Analiza las razones de algunos educadores y académicos en el área de humanidades que han adoptado una actitud hostil hacia la ciencia.

Scientific American. Science versus antiscience. (Special report). Enero 1997, 96-101. Presenta tres diferentes movimientos anticiencia: creacionista, feminista y medios de comunicación.

Smith, B. (1995). Formal ontology, common sense and cognitive science. *International Journal of Human-Computer Studies, 43*(5-6), 641-667. Un artículo que examina el sentido común y la ciencia cognitiva.

Timpane, J. (1995). How to convince a reluctant scientist. *Scientific American, 272,* 104. Este artículo advierte que demasiada originalidad en la ciencia puede conducir a la falta de aceptación y a la dificultad en su comprensión. También analiza cómo la aceptación científica está gobernada tanto por información previa como nueva, y por la reputación del científico.



CAPÍTULO 2

PROBLEMAS E HIPÓTESIS

- PROBLEMAS
- CRITERIOS DE LOS PROBLEMAS Y ENUNCIADOS DE PROBLEMAS
- HIPÓTESIS
- IMPORTANCIA DE LOS PROBLEMAS E HIPÓTESIS
- VIRTUDES DE LOS PROBLEMAS E HIPÓTESIS
- PROBLEMAS, VALORES Y DEFINICIONES
- GENERALIDAD Y ESPECIFICIDAD DE LOS PROBLEMAS E HIPÓTESIS
- LA NATURALEZA MULTIVARIABLE DE LA INVESTIGACIÓN Y PROBLEMAS DEL COMPORTAMIENTO
- COMENTARIOS FINALES: EL PODER ESPECIAL DE LAS HIPÓTESIS

Mucha gente cree que la ciencia es en lo fundamental una actividad de recolección de hechos. M.R. Cohen (1956/1997, p. 148) lo planteó de otra forma:

> No hay un progreso genuino en el discernimiento científico a través del método baconiano de acumular hechos empíricos sin una hipótesis o anticipación de la naturaleza. Sin alguna idea que nos guíe no sabemos qué hechos recolectar... no podemos determinar qué es y qué no es relevante.

Las personas sin formación científica con frecuencia tienen la idea que el científico es un individuo objetivo en extremo que recolecta datos sin tener ideas preconcebidas. Poincare (1952/1996, p. 143) señaló lo equivocado de esta idea: "Se dice a menudo que los experimentos deben realizarse sin ideas preconcebidas. Eso es imposible. No sólo haría que todo experimento fuera improductivo, sino que, aun deseándolo, no podríamos hacerlos."

Problemas

No siempre le es posible al investigador definir el problema de una manera simple, clara y completa. A menudo, puede tener una noción general, difusa e, inclusive, confusa del pro-

blema. Esto se debe a la naturaleza compleja de la investigación científica. Es posible incluso que, pueda tomarle años de exploración, reflexión e investigación el poder definir una pregunta de forma clara. Sin embargo, enunciar de manera adecuada el problema de investigación es una de las partes fundamentales del proceso. La dificultad para enunciar un problema de investigación de forma satisfactoria en un momento dado no debe hacernos perder de vista lo necesario y deseable que resulta.

Con esta dificultad en mente, podemos establecer un principio fundamental: si queremos resolver un problema, en general debemos conocerlo. Se puede decir que gran parte de la solución estriba en conocer lo que se trata de hacer. Otra parte está en entender qué es un problema y, en especial, un problema científico.

¿Qué constituye un buen enunciado del problema? Aunque los problemas de investigación difieren en gran medida y no existe una fórmula "correcta" para enunciar problemas, es posible aprender y utilizar para nuestro beneficio ciertas características de los problemas y de los enunciados de problemas. Para empezar, consideremos dos o tres ejemplos de problemas de investigación publicados y estudiemos sus características. Primero, tomemos el problema del estudio realizado por Hurlock (1925),[1] mencionado en el capítulo 1: ¿Cuáles son los efectos de diferentes tipos de incentivos en el rendimiento del alumno? Observe que el problema está enunciado en forma de pregunta. En este campo, la forma más simple es la mejor. También conviene señalar que el problema establece una relación entre variables, en este caso, entre las variables *incentivos* y *rendimiento del alumno* (logro). (El término *variable* será definido de manera formal en el capítulo 3. Por ahora, se usará para nombrar un fenómeno o un constructo, que asume un conjunto de diferentes valores numéricos.)

Un *problema*, entonces, es un enunciado u oración interrogativa que pregunta: ¿qué relación existe entre dos o más variables? La respuesta constituye aquello que se busca en la investigación. Un problema, en la mayoría de los casos, tendrá dos o más variables. En el ejemplo de Hurlock, el enunciado del problema relaciona incentivos con rendimiento del alumno. Otro problema, estudiado en el experimento clásico de Bahrick (1984, 1992) está asociado con preguntas de edad y vejez: ¿Cuánto de lo que ahora estás estudiando recordarás dentro de diez años? ¿Cuánto de esto mismo podrás recordar dentro de cincuenta años? ¿Cuánto recordarás después, si nunca lo utilizas? La pregunta formal de Bahrick es: ¿la memoria semántica involucra procesos separados? Una variable es la cantidad de tiempo que transcurre desde que el material se aprendió por primera vez; la segunda podría ser la calidad del aprendizaje original; y la otra variable es el recuerdo (u olvido). Veamos otro problema de Little, Sterling y Tingstrom (1996), que es muy diferente: ¿Influyen las claves geográficas y las características raciales en la atribución (culpa percibida)? Una variable son las claves geográficas; la segunda sería la información racial, y la tercera, la atribución.

No todos los problemas de investigación contienen dos o más variables claras. Por ejemplo, en psicología experimental, el foco de la investigación con frecuencia está en procesos psicológicos como la memoria y la categorización. Rosch (1973) en su influyente y justificadamente bien conocido estudio de categorías perceptuales hizo la siguiente pregunta: ¿Existen categorías no arbitrarias ("naturales") de color y forma? Aunque la relación entre dos o más variables no es aparente en este enunciado del problema, en la investigación real las categorías estaban relacionadas con el aprendizaje. Hacia el final de este libro se verá que los problemas de investigación factorial analítica también carecen

[1] Cuando referimos problemas e hipótesis de la literatura, no siempre usamos las palabras de los autores. De hecho, los enunciados de muchos de los problemas son nuestros y no de los autores citados. Algunos autores sólo usan enunciados de problemas, algunos sólo hipótesis, y otros usan ambos.

de la forma de las relaciones antes planteada. Sin embargo, en la mayoría de los problemas de investigación del comportamiento, se estudian las relaciones entre dos o más variables; por ello, enfatizaremos ese tipo de enunciados de relación.

Criterios de los problemas y enunciados de problemas

Existen tres criterios de buenos problemas y enunciados de problemas. El primero: el problema debe expresar una relación entre dos o más variables. En efecto plantea preguntas como la siguiente: ¿*A* está relacionada con *B*? ¿Cómo están relacionadas *A* y *B* con *C*? ¿Cómo está *A* relacionada con *B* bajo las condiciones *C* y *D*? La excepción a esta consideración ocurre casi siempre en investigación metodológica o taxonómica.

Segundo: el problema debe ser enunciado de manera clara y sin ambigüedades en forma de pregunta. En lugar de decir, por ejemplo "el problema es..." o "El propósito de este estudio es..." resulta necesario plantear una pregunta. Las preguntas tienen la virtud de presentar los problemas directamente. El propósito de un estudio no es por fuerza el mismo que el problema de un estudio. El propósito del estudio de Hurlock, por ejemplo, fue arrojar luz sobre el uso de incentivos en las situaciones escolares. El problema consistió en la pregunta acerca de la relación entre incentivos y rendimiento. Otra vez, mientras más simple, mejor: formule una pregunta.

El tercer criterio con frecuencia es difícil de satisfacer. Demanda que el problema y su enunciado *impliquen* la posibilidad de ser sometidos a una prueba empírica. Un problema que no contenga implicaciones para probar las relaciones que enuncia, no constituye un problema científico. Esto significa no sólo que se enuncie una relación real, sino también que las variables de la relación puedan ser medidas de alguna forma. Hay muchas preguntas interesantes e importantes que no constituyen preguntas científicas tan sólo porque no son susceptibles de prueba. Ciertas preguntas filosóficas y teológicas, aunque importantes para quienes las consideran, no pueden ser probadas empíricamente, por lo que no generan interés para el científico como tal. La pregunta epistemológica "¿Cómo conocemos?" es una pregunta de ese tipo. La educación plantea muchas preguntas interesantes pero no científicas, por ejemplo "¿Mejora la educación democrática el aprendizaje de los jóvenes?" "¿Son buenos los procesos grupales para los niños?" Estas preguntas pueden ser etiquetadas como metafísicas en el sentido en que están, al menos así enunciadas, fuera de la posibilidad de una prueba empírica. Las principales dificultades estriban en que algunas no constituyen relaciones, y es muy difícil o imposible definir la mayoría de sus constructos de forma que puedan ser medidos.

Hipótesis

Una *hipótesis* es un enunciado conjetural de la relación entre dos o más variables. Las hipótesis siempre se presentan en forma de enunciados declarativos y relacionan, de manera general o específica, variables con variables. Hay dos criterios que definen a las "buenas" hipótesis y a sus enunciados. Son los mismos que mencionamos para los problemas y sus enunciados. 1) Las hipótesis son enunciados acerca de las relaciones entre variables. 2) Las hipótesis contienen implicaciones claras para probar las relaciones enunciadas. Estos criterios significan que los enunciados de hipótesis contienen dos o más variables, que son medibles o pueden serlo, y que especifican cómo están relacionadas las variables.

Permítanos mencionar tres hipótesis de la literatura y aplicarles estos criterios. La primera hipótesis procede de un estudio de Wegner y colaboradores. (1987) que parece desafiar el sentido común: a mayor supresión de pensamientos indeseados, mayor preocu-

pación por ellos (represión ahora; obsesión más tarde). Aquí se establece una relación entre una variable, supresión de una idea o pensamiento, y otra variable, preocupación u obsesión. Dado que ambas se definen y miden con facilidad, las implicaciones para probar la hipótesis también se conciben sin esfuerzo. Los criterios están satisfechos. En el estudio de Wegner y colaboradores, se les pidió a los sujetos que *no* pensaran en un "oso blanco". Cada vez que pensaran en él, debían de tocar una campana. El número de campanadas indicaba el nivel de preocupación. Una segunda hipótesis, que resulta inusual y corresponde al estudio de Ayres y Hughes (1986), enuncia la relación de una forma que llamamos nula: el nivel de ruido o música no tiene efecto en el funcionamiento visual. La relación se establece con claridad: una variable, intensidad del sonido (por ejemplo, música), se relaciona con otra, funcionamiento visual, a través de las palabras "no tiene efecto en". En el criterio de potencia de ser probada, sin embargo, encontramos dificultades. Nos enfrentamos con el problema de definir "funcionamiento visual" e "intensidad" de forma que puedan medirse. Si podemos resolver este problema de manera satisfactoria, entonces, tenemos en definitiva una hipótesis. Ayres y Hughes lo hicieron al definir intensidad como 107 decibeles y funcionamiento visual en términos de una puntuación en una tarea de agudeza visual. Esta hipótesis permitió contestar una pregunta que la gente con mucha frecuencia se hace: "¿por qué bajamos el volumen del estéreo del auto cuando *buscamos* una dirección?". Ayres y Hughes encontraron una caída marcada en el funcionamiento perceptual cuando el nivel de música llegaba a 107 decibeles.

La tercera hipótesis representa una categoría numerosa e importante. En ella la relación es indirecta, oculta. En general enuncia que los grupos *A* y *B* diferirán en alguna característica. Por ejemplo: las mujeres creen, con mayor frecuencia que los hombres, que deben perder peso aun cuando éste se encuentre dentro de los límites normales (Fallon y Rozin, 1985). Esto es, que las mujeres difieren de los hombres en cuanto a la percepción de su figura corporal. Observe que este enunciado está un paso más allá de la hipótesis real que puede plantearse como: la percepción de la figura corporal es, en parte, una función del género. Si los enunciados posteriores constituyeran la hipótesis enunciada, entonces la primera podría llamarse una subhipótesis o una predicción específica basada en la hipótesis original.

Consideremos otra hipótesis de este tipo pero dando un paso más adelante. Los individuos que tienen características iguales o similares tendrán actitudes similares hacia objetos cognitivos significativamente relacionados con su papel ocupacional (Saal y Moore, 1993). (*Los objetos cognitivos* se definen como algo concreto o abstracto, percibido y "conocido" por los individuos. Personas, grupos, ascenso en el trabajo o en las calificaciones, el gobierno y la educación son algunos ejemplos.) La relación en este caso es, por supuesto, entre características personales y actitudes (hacia un objeto cognitivo relacionado con la característica personal, por ejemplo, género y actitudes hacia otros que reciben una promoción). Para probar esta hipótesis, sería necesario tener al menos dos grupos, cada uno con una característica diferente, y después comparar las actitudes de ambos grupos. Por ejemplo, como en el caso del estudio de Saal y Moore, la comparación sería entre hombres y mujeres. Serían comparados en relación a su evaluación hacia el ascenso dado a un compañero de trabajo del mismo sexo o del opuesto. En este ejemplo, se satisfacen los criterios.

Importancia de los problemas e hipótesis

Hay poca duda de que las hipótesis son herramientas importantes e indispensables de la investigación científica. Existen tres razones principales para esta creencia. La primera es que son, digamos, los instrumentos de trabajo de la teoría. Las hipótesis pueden deducirse

a partir de la teoría y de otras hipótesis. Si por ejemplo, trabajamos en una teoría sobre la agresividad, se presume que buscamos causas y efectos del comportamiento agresivo. Es posible que hayamos observado casos de comportamiento agresivo ocurrido después de circunstancias frustrantes. La teoría, entonces, puede incluir la proposición: la frustración produce agresividad (Berkowitz, 1983; Dill y Anderson, 1995; Dollard, Doob, Miller, Mowrer, y Sears, 1939). A partir de esta amplia hipótesis, podemos deducir hipótesis más específicas, como por ejemplo: impedir que los niños alcancen sus metas deseadas (frustración) generará pleitos entre ellos (agresión); si los niños son privados del amor paterno (frustración), reaccionarán en parte, con un comportamiento agresivo.

La segunda razón es que es posible someter a prueba las hipótesis y demostrar que son probablemente verdaderas o probablemente falsas. No se prueban hechos aislados, como se dijo antes, sólo relaciones. Es probable que la principal razón de usar hipótesis en la investigación científica sea que constituyen proposiciones relacionales. En esencia, son predicciones del tipo: "si *A*, entonces *B*", que utilizamos para probar la relación entre *A* y *B*. Dejamos que los hechos tengan la oportunidad de establecer la probable veracidad o falsedad de la hipótesis.

La tercera razón es que las hipótesis son herramientas poderosas para el avance del conocimiento porque permiten al científico ir más allá de sí mismo. Aunque desarrolladas por humanos, las hipótesis existen, pueden ser probadas y puede demostrarse que son probablemente correctas o incorrectas de manera independiente a los valores y opiniones de una persona (sesgos). Esto resulta crítico: no habría ciencia, en sentido completo alguno, sin las hipótesis.

Tan importantes como las hipótesis son los problemas tras ellas. Como Dewey (1938/1982, pp. 105-107) ha señalado, la investigación por lo general empieza con un problema. Indica que primero hay una situación indeterminada en la que las ideas son vagas, aparecen dudas, y el pensador queda perplejo. Agrega que el problema no se enuncia; de hecho, no puede ser enunciado hasta que uno ha experimentado una situación tan indeterminante.

La indeterminación, sin embargo, deberá, en última instancia ser eliminada. Aunque es cierto, como se señaló antes, que el investigador con frecuencia puede tener sólo una noción general y difusa del problema, tarde o temprano habrá de definir una idea clara de lo que el problema es. Aunque este enunciado parezca obvio, una de las cosas más difíciles de lograr, es enunciar de una manera clara y completa el problema de investigación. En otras palabras, uno debe saber qué es lo que trata encontrar. Cuando por fin se identifica, el problema ya está en camino a la solución.

Virtudes de los problemas e hipótesis

Los problemas y las hipótesis tienen virtudes importantes: 1) dirigen la investigación (las relaciones expresadas en las hipótesis indican al investigador lo que debe hacer); 2) los problemas e hipótesis, dado que son de ordinario enunciados relacionales generalizados, permiten al investigador deducir manifestaciones empíricas específicas implicadas en ellos. Podemos decir, de acuerdo con Guida y Ludlow (1989): si es un hecho verdadero que los niños de un tipo de cultura (Chile) tienen un mayor grado de ansiedad que los niños de otro tipo de cultura (blancos estadounidenses), entonces los niños en la cultura chilena deben rendir menos en lo académico que los niños en la cultura estadounidense. Los niños chilenos quizá también debieran presentar una menor autoestima o un *locus* de control más externo en lo que se refiere a la escuela y a la labor académica.

Hay diferencias importantes entre problemas e hipótesis. Las hipótesis, si están enunciadas de manera apropiada, pueden ser probadas. Una hipótesis dada puede ser demasia-

do amplia para ser probada de forma directa; sin embargo, si es una "buena" hipótesis, es posible deducir a partir de ella otras que sí lo sean. Los hechos o las variables no se prueban como tales. Se prueban las relaciones enunciadas por las hipótesis. Un problema no puede ser resuelto de manera científica a menos que se reduzca a su forma de hipótesis, ya que un problema es una pregunta, generalmente de naturaleza amplia, que no puede probarse en forma directa. No se someten a prueba preguntas como: ¿la presencia o ausencia de otra persona en un sanitario público afecta la higiene personal? (Pedersen, Keithly y Brady, 1986). ¿Las sesiones de consejería grupal disminuyen el nivel de morbilidad psiquiátrica en oficiales de policía? (Doctor, Cutris e Issacs, 1994). Quizás uno probaría una o más hipótesis deducidas de estas preguntas. Por ejemplo, para estudiar el último problema, uno puede hipotetizar que los oficiales de policía que asisten a sesiones de consejería para reducir el estrés requerirán menos días de incapacidad por enfermedad que aquéllos que no asisten. La hipótesis para el primer problema podría indicar que la presencia de una persona en un sanitario público hará que otras se laven las manos.

Los problemas e hipótesis permiten que avance el conocimiento científico al ayudar al investigador a confirmar o refutar una teoría. Suponga que un investigador en el campo de la psicología aplica a unos sujetos tres o cuatro pruebas, entre las cuales hay una para evaluar la ansiedad relacionada con una prueba aritmética. Al calcular de manera rutinaria las correlaciones entre las tres o cuatro pruebas, uno encuentra que la correlación entre ansiedad y aritmética es negativa. De lo anterior se deduce que a mayor ansiedad, menor puntuación en la prueba de aritmética. Sin embargo, es muy probable que esta relación sea fortuita e incluso espuria, pero si el investigador hubiera hipotetizado la relación con base en una teoría, tendría mayor confianza en sus resultados. El investigador que no hipotetiza relaciones en forma previa, no permite que los hechos prueben o rechacen nada. Las palabras *probar* y *rechazar* no deben tomarse en su sentido literal: una hipótesis nunca se prueba o refuta realmente. Para ser más precisos, deberíamos decir algo del tipo de: el peso de la evidencia está del lado de la hipótesis o el peso de la evidencia arroja dudas sobre la hipótesis. Braithwaite (1953/1996, p. 14) dice:

> De este modo, la evidencia empírica nunca prueba la hipótesis: en casos apropiados podemos decir que *se establece* (itálicas agregadas) la hipótesis, lo que significa que la evidencia hace que sea razonable aceptar la hipótesis; pero ésta nunca *prueba* la hipótesis en el sentido de que la hipótesis sea una consecuencia lógica de la evidencia.

Este uso de la hipótesis es similar a participar en un juego de azar. Se establecen las reglas del juego y se definen las apuestas por adelantado. Uno no puede cambiar las reglas después de un resultado, como tampoco se pueden cambiar las apuestas una vez hechas. Uno no tira los dados primero y luego apuesta. No sería "justo". De igual forma, si en primera instancia se recolectan datos y después se toma uno de ellos y se llega a una conclusión con base en él, se han violado las reglas del juego científico. El juego no es "justo" porque el investigador puede capitalizar fácilmente, digamos, dos relaciones importantes de las cinco a prueba. Las otras tres, por lo general, se olvidan. En un juego "justo", se cuenta cada tiro del dado, en el sentido de que se gana o no con base en el resultado de cada tirada.

Las hipótesis dirigen la investigación. Como Darwin señaló hace más de 100 años, todas las observaciones han de ser a favor o en contra de algún punto de vista para tener alguna utilidad. Las hipótesis incorporan aspectos de la teoría bajo prueba de forma susceptible o casi susceptible de ser probada. Antes se dio un ejemplo de la teoría del reforzamiento en el que se dedujeron hipótesis demostrables a partir del problema general. Podemos demostrar la importancia del reconocimiento de esta función de las hipótesis al introducirnos por la puerta trasera y usar una teoría muy difícil o quizás imposible de probar. La teoría de Freud de la ansiedad incluye el constructo de la represión. Con este

término Freud se refería a la introducción forzada de ideas inaceptables en lo profundo del inconsciente. Para probar la teoría freudiana de la ansiedad es necesario deducir relaciones sugeridas por la teoría. Estas deducciones por fuerza deberán incluir el concepto de represión que implica el constructo del inconsciente. Es posible formular hipótesis que utilizan estos constructos; para probar la teoría han de ser formuladas de esta manera. Pero probarlos resulta más difícil debido a la extrema dificultad para definir términos como "represión" e "inconsciente" de manera que puedan medirse. Hasta hoy nadie ha tenido éxito al definir estos dos constructos sin apartarse en gran medida del significado y uso freudianos originales. Las hipótesis constituyen, entonces, puentes importantes entre la teoría y la investigación empírica.

Problemas, valores y definiciones

Para establecer con claridad la naturaleza de los problemas y de las hipótesis, analizaremos ahora dos o tres errores comunes. En primer término, los problemas científicos no son preguntas morales ni éticas: ¿Son las medidas disciplinarias de tipo punitivo perjudiciales para los niños? ¿Debiera ser el liderazgo de una organización de tipo democrático? ¿Cuál es la mejor forma de enseñar a los estudiantes universitarios? Formular estas preguntas equivale a presentar cuestionamientos de valor y juicio que la ciencia no puede contestar. Muchas de las que se han llamado hipótesis no lo son en absoluto. Por ejemplo: el método de enseñanza a pequeños grupos es mejor que el método expositivo. Éste es un enunciado de valor; constituye un acto de fe, no una hipótesis. Si fuera posible establecer una relación entre las variables, y definirlas de manera que se pudiera probar esa relación, entonces podríamos contar con una hipótesis. Pero no hay una forma científica de someter a prueba preguntas de valor.

Una forma rápida y relativamente fácil de detectar preguntas y enunciados de valor consiste en buscar palabras como *debe*, *debería*, *mejor que* (en lugar de *mayor que*), así como palabras similares que indiquen juicios culturales o personales, o preferencias (sesgos). Sin embargo, los enunciados de valor son engañosos. Aunque resulta obvio que un enunciado que incluye la palabra "debería" es un enunciado de valor, otros tipos no son tan evidentes. Consideremos el enunciado: los métodos autoritarios de enseñanza conducen a un pobre aprendizaje. En este caso sí hay una relación. Pero el enunciado falla como una hipótesis científica en tanto que incorpora dos expresiones de valor: "métodos autoritarios de enseñanza" y "pobre aprendizaje", ninguna de las cuales puede definirse con propósitos de medición sin borrar las palabras *autoritario* y *pobre*.[2]

Con frecuencia se formula otra clase de enunciados que no constituyen hipótesis o que son hipótesis pobres, en especial en el campo de la educación. Consideremos, por ejemplo, el siguiente: los cursos de tronco común representan una experiencia enriquecedora. Otro tipo de enunciado que se usa con frecuencia, es la generalización vaga: es posible identificar las habilidades de lectura en el segundo grado; La meta del individuo auténtico es la autorrealización; El prejuicio se relaciona con ciertos rasgos de personalidad.

Otro defecto común de los enunciados de problema aparece a menudo en las tesis doctorales: enlistar aspectos metodológicos o "problemas" como subproblemas. Estos aspectos metodológicos poseen dos características que hacen fácil detectarlos: 1) No son

[2] Un caso ya casi clásico del uso de la palabra *autoritario* es la frase que a veces se escucha entre educadores: el método expositivo es autoritario. Esto parece indicar que quien lo dice no gusta del método expositivo y lo considera negativo. De forma similar, una de las formas más efectivas de criticar a un maestro es decir que es autoritario.

problemas sustantivos que surgen del problema básico; y 2) Se relacionan con técnicas o métodos de muestreo, medición o análisis. En general, no aparecen en forma de pregunta y contienen palabras tales como *probar, determinar, medir*: "para *determinar* la confiabilidad de los instrumentos usados en esta investigación;" "para *probar* la significación de las diferencias entre las medias" o "para *asignar* alumnos de manera aleatorizada a los grupos experimentales", son ejemplos de esta noción equivocada de problemas y subproblemas.

Generalidad y especificidad de los problemas e hipótesis

Una dificultad que el investigador por lo general encuentra y que casi todos los estudiantes que trabajan en una tesis hallan molesta, es la generalidad y la especificidad de los problemas e hipótesis. Si el problema es muy general, es demasiado vago para ser sometido a prueba. Así, desde el punto de vista científico resulta inútil, aunque puede ser interesante para leer. Los problemas e hipótesis demasiado generales o vagos son comunes. Por ejemplo: La creatividad es una función de la autorrealización del individuo; La educación democrática potencia el aprendizaje social y la forma cívica; El autoritarismo en el salón de clases universitario inhibe la imaginación creativa del estudiante. Todos resultan problemas interesantes, pero en su forma actual son por completo inútiles en el terreno científico, en tanto que no pueden ser sometidos a prueba y porque parecen sugerir una seguridad espuria de que constituyen hipótesis que "algún día" pueden ser probadas.

Términos tales como "creatividad", "autorrealización", "democracia" y "autoritarismo" no tienen, al menos hasta ahora, referentes empíricos adecuados.[3] Es cierto que podemos definir *creatividad*, en una forma limitada al especificar una o dos pruebas de creatividad. Éste puede ser un procedimiento legítimo; sin embargo, al emplearlo, corremos el riesgo de alejarnos del término original y de su significado. Esto es en particular cierto cuando hablamos de creatividad artística. Desde luego, con frecuencia aceptamos el riesgo con tal de investigar problemas importantes. Aun así, términos como "democracia" son casi imposibles de definir. Incluso cuando lo hacemos, a menudo descubrimos que hemos destruido su significado original. Una excepción sobresaliente es la definición y la medición de "democracia" de Bollen (1980). Examinaremos ambas en otros capítulos.

El otro extremo es caer en demasiada especificidad. Todo estudiante ha escuchado que es necesario reducir los problemas a una dimensión manejable. Esto es cierto, pero por desgracia, podemos reducirlo tanto hasta hacerlo desaparecer. En general, mientras más específicos son el problema o la hipótesis, más claras resultan sus implicaciones a probar. Sin embargo, el precio que podemos pagar es la trivialidad. Los investigadores no pueden manejar problemas demasiado amplios por su tendencia a ser demasiado vagos en cuanto a las operaciones adecuadas de investigación. Por otro lado, en su entusiasmo por reducir el problema a un tamaño manejable o por encontrar un problema manipulable, pueden terminar con su vida, y convertirlo en trivial o carente de importancia. Por ejemplo, una tesis sobre la simple relación entre velocidad de lectura y tamaño de la letra, por muy interesante e importante que pudiera parecer, resulta débil para un estudio doctoral. El estudiante de ese nivel necesitará ampliar el tema al recomendar una comparación entre géneros y considerar variables como cultura y antecedentes familiares. El investigador podría también expandir el estudio para concentrarse en los niveles de iluminación y

[3] Aunque se ha conducido con éxito una variedad de estudios sobre autoritarismo, no es claro que entendamos lo que significa autoritarismo en el salón de clases. Por ejemplo, una acción de un maestro autoritario en un salón de clases puede no serlo en otra aula. La mencionada conducta democrática exhibida por un maestro puede ser etiquetada como autoritarismo si la muestra otro docente. Tal elasticidad no pertenece a la ciencia.

el tipo de letra. Demasiada especificidad quizás sea más dañina que demasiada generalidad. El investigador puede estar en posición de contestar una pregunta específica pero no podrá generalizar los hallazgos a otras situaciones o grupos de personas. A cualquier precio, alguna clase de compromiso debe establecerse entre generalidad y especificidad. La capacidad para definir tal compromiso de manera efectiva es, en parte, función de la experiencia, y en parte, del estudio crítico de los problemas de investigación.

He aquí algunos ejemplos de problemas de investigación contrastantes, ya sea muy generales o muy específicos:

1)	Demasiado general:	Existen diferencias de género al jugar.
	Demasiado específico:	La puntuación de Carlos será 10 puntos mayor que la de Carol en el juego profesional Tetris.
	Cerca del ideal:	Habrá mayor transferencia de aprendizaje al practicar videojuegos con niños que con niñas.
2)	Demasiado general:	Las personas pueden leer letras de mayor tamaño más rápido que las letras más pequeñas.
	Demasiado específica:	Los alumnos de último año de la escuela Duarte pueden leer tipos de letra de 24 puntos más rápido que tipos de letra de 12 puntos.
	Cerca del ideal:	Una comparación de tres diferentes tamaños de letra y agudeza visual en la velocidad de lectura y de comprensión.

La naturaleza multivariable de la investigación y problemas del comportamiento

Hasta este punto, la discusión de problemas e hipótesis se ha limitado a dos variables, x y y. Debemos corregir cualquier impresión de que tales problemas e hipótesis son la norma en la investigación del comportamiento. Los investigadores en psicología, sociología, educación y otras ciencias del comportamiento se han concientizado de la naturaleza multivariable de este tipo de estudios. En lugar de decir: si p, entonces q, es frecuente y más apropiado decir: si $p_1, p_2 \ldots, p_n$, entonces q; o bien: si p, entonces q, bajo las condiciones r, s y t.

A continuación, un ejemplo que puede aclarar este punto. En lugar de simplemente formular la hipótesis: si hay frustración, entonces hay agresividad, es más realista reconocer la naturaleza multivariable de los determinantes e influencias de la agresividad. Esto se logra al decir, por ejemplo: si se es muy inteligente, de clase media, varón y frustrado, entonces hay agresividad; o bien: si hay frustración, entonces hay agresividad bajo las condiciones de alta inteligencia, clase media y sexo masculino. En lugar de tener una x, nosotros ahora tenemos cuatro x. Aunque un fenómeno puede ser el más importante para determinar o influir en otro fenómeno, es poco probable que la mayoría de los fenómenos de interés para los científicos del comportamiento sean determinados de forma simple. Es mucho más probable que lo sean de manera múltiple. Es mucho más probable que la agresividad sea el resultado de diversas influencias que actúan de forma compleja. Más aún, la agresividad en sí misma contiene múltiples aspectos. Después de todo, hay diferentes clases de agresividad.

Los problemas y las hipótesis, entonces deben reflejar la complejidad multivariable de la realidad psicológica, sociológica y educativa. Hablaremos de una x y una y, en particular en la parte inicial de este libro. Sin embargo, *es preciso* entender que la investigación del

comportamiento, que tuvo un enfoque casi exclusivamente univariado, se ha tornado cada vez más multivariable. Nos hemos propuesto usar la palabra "multivariable" en lugar de "multivariada" por una razón importante. De forma tradicional los estudios "multivariados" son aquéllos que tienen más de una variable *y* y una o más variables *x*. Cuando hablamos de una variable *y* y más de una variable *x* se usa el término "multivariable" que resulta más apropiado para hacer la distinción. Por ahora usaremos "univariado" para indicar una *x* y una *y*. De manera estricta, el término "univariado" también se aplica a *y*. Pronto encontraremos conceptos y problemas de naturaleza multivariada. Secciones posteriores del libro estarán enfocadas en especial a un enfoque y énfasis de este tipo. Para más explicaciones sobre las diferencias entre *multivariable* y *multivariado* (véase Kleinbaum, Kupper, Muller y Nizam, 1997).

Comentarios finales: el poder especial de las hipótesis

A veces se oye decir que las hipótesis son innecesarias en la investigación. Algunos sienten que las hipótesis restringen innecesariamente su imaginación investigadora y que el trabajo de la ciencia y de la investigación científica es descubrir cosas nuevas y no elaborar lo obvio. Algunos piensan que las hipótesis son obsoletas. Tales afirmaciones resultan engañosas y malinterpretan el propósito de las hipótesis.

Casi puede decirse que las hipótesis son uno de los instrumentos más poderosos que ha inventado el hombre para alcanzar un conocimiento confiable. Observamos un fenómeno; especulamos sobre sus causas posibles. Naturalmente, nuestra cultura tiene respuestas para dar cuenta de la mayoría de los fenómenos —muchas correctas, muchas incorrectas, muchas una mezcla de hechos y supersticiones, y muchas pura superstición—. Es obligación del científico dudar de la mayor parte de las explicaciones sobre los fenómenos. Esas dudas son sistémicas. Los científicos insisten en someter las explicaciones sobre los fenómenos a una prueba empírica controlada. Para lograrlo, formulan las explicaciones en términos de teorías e hipótesis. De hecho, las explicaciones constituyen hipótesis. Los científicos sólo disciplinan la cuestión al escribirla en forma de hipótesis sistemáticas y comprobables. Si una explicación no puede formularse en términos de una hipótesis comprobable, deberá considerarse como una explicación metafísica y, por lo tanto, no susceptible de investigación científica. Como tal, los científicos la rechazan como carente de interés.

El poder de las hipótesis va más allá. La hipótesis constituye una predicción: indica que si ocurre *x*, también ocurrirá *y*; esto es, *y* se predice a partir de *x*. Entonces si se hace que *x* ocurra (es decir, que varíe) y se observa que *y* también ocurre (o sea, varía de forma concomitante), entonces la hipótesis se confirma. Resulta una evidencia más poderosa que la simple observación, sin predicción, la covariación de *x* y *y*. Es más poderosa en el sentido de apuesta-juego antes discutido. El científico apuesta a que *x* conduce hacia *y*. Si en un experimento, *x* conduce en efecto a *y*, entonces habrá ganado la apuesta. Una persona no puede sólo entrar en el juego en cualquier momento y observar una ocurrencia común quizá fortuita de *x* y *y*. No se juega de esta forma (al menos no en nuestra cultura). La persona debe jugar de acuerdo con las reglas, y las reglas en la ciencia están hechas para minimizar el error y la falibilidad. Las hipótesis forman parte de las reglas del juego científico.

Aun cuando no se confirmen las hipótesis, tienen poder. Aun cuando *y* no covaríe con *x*, el conocimiento avanza. Los hallazgos negativos en ocasiones resultan tan importantes como los positivos, puesto que reducen el universo total de la ignorancia, y algunas veces señalan hacia otras hipótesis y líneas de investigación. *Pero el científico no puede distinguir la evidencia positiva de la negativa hasta usar una hipótesis.* Por supuesto, es posible conducir

una investigación sin hipótesis, en particular en el caso de estudios exploratorios, pero es difícil concebir a la ciencia moderna en toda su rigurosa y disciplinada fertilidad sin la guía y poder de las hipótesis.

RESUMEN DEL CAPÍTULO

1. Formular un problema de investigación no es una tarea fácil. El investigador empieza con una noción general difusa y vaga que gradualmente se refina. Los problemas de investigación varían en gran medida y no existe un único camino correcto para enunciar el problema.
2. Tres criterios de problemas y enunciados de problema adecuados son:
   a) El problema debe expresarse como una relación entre dos o más variables.
   b) El problema debe ser redactado en forma de pregunta.
   c) El enunciado del problema debe implicar la posibilidad de ser sometido a una prueba empírica.
3. Una hipótesis es un enunciado conjetural de la relación entre dos o más variables. Ésta se redacta en forma de enunciado declarativo. Los criterios para una hipótesis apropiada son los mismos que los usados para los problemas, que se señalan en el punto anterior.
4. La importancia de los problemas e hipótesis es que:
   a) Constituyen un instrumento de trabajo de la ciencia y un enunciado de trabajo específico de la teoría.
   b) Las hipótesis pueden ser sometidas a prueba y ser predictivas.
   c) Contribuyen al avance del conocimiento.
5. Las virtudes de los problemas y de las hipótesis son:
   a) Dirigen la investigación.
   b) Permiten al investigador deducir manifestaciones empíricas específicas.
   c) Sirven como puente entre teoría e investigación empírica.
6. Los problemas científicos no constituyen preguntas éticas y morales. La ciencia no puede contestar preguntas de valor o de juicio.
7. Para detectar preguntas de valor es necesario buscar palabras tales como *mejor que*, *debería*, o *habría que*.
8. Otro defecto común de los enunciados de problema es enlistar aspectos metodológicos como subproblemas. La falla consiste en que:
   a) No son problemas sustantivos que provengan del problema básico en forma directa.
   b) Están relacionados con técnicas o métodos de muestreo, medición, o análisis; no se presentan en forma de pregunta.
9. En cuanto a los problemas, es necesario establecer un equilibrio para que no sea ni demasiado general ni demasiado específico. Esta habilidad se desarrolla con la experiencia.
10. Los problemas e hipótesis deben reflejar la complejidad multivariada de la realidad del ámbito de las ciencias del comportamiento.
11. La hipótesis representa uno de los más poderosos instrumentos inventados para obtener conocimiento confiable. Tiene la capacidad de ser predictiva. Un hallazgo negativo para una hipótesis puede servir para eliminar una posible explicación y generar otras hipótesis y líneas de investigación.

SUGERENCIAS DE ESTUDIO

1. Utilice los siguientes nombres de variables para redactar problemas de investigación e hipótesis: frustración, logro académico, inteligencia, habilidad verbal, raza, clase social (estatus socioeconómico), sexo, reforzamiento, métodos de enseñanza, elección ocupacional, conservadurismo, educación, ingresos, autoridad, necesidad de logro, cohesión de grupo, obediencia, prestigio social, permisividad.
2. A continuación se presentan diez problemas de investigación tomados de la literatura. Estúdielos con cuidado, elija dos o tres y construya hipótesis con base en ellos.
   a) ¿Tienen diferentes puntuaciones en una prueba de ansiedad los niños de diferentes grupos étnicos? (Guida y Ludlow, 1989)
   b) ¿Las situaciones de cooperación social conducen a mayores niveles de motivación intrínseca? (Hom, Berger, Duncan, Miller y Belvin, 1994)
   c) ¿Las expresiones faciales de las personas influyen en las respuestas afectivas? (Strack, Martin y Stepper, 1988)
   d) ¿Respetarán los jurados las instrucciones e indicaciones judiciales prohibitivas? (Shaw y Skolnick, 1995)
   e) ¿Cuáles son los efectos positivos del uso de cojines de presión alternante para prevenir llagas en pacientes terminales atendidos en casa? (Stoneberg, Pitcock y Myton, 1986)
   f) ¿Cuáles son los efectos del condicionamiento pavloviano temprano en el condicionamiento pavloviano tardío? (Lariviere y Spear, 1996)
   g) ¿Depende la eficacia de la codificación de información en la memoria de largo plazo de lo novedoso que ésta sea? (Tulving y Kroll, 1995)
   h) ¿Cuál es el efecto del consumo de alcohol en la probabilidad de uso del condón durante el sexo ocasional? (MacDonald, Zanna y Fong, 1996)
   i) ¿Hay diferencias por género para predecir las decisiones relativas al retiro? (Talaga y Beehr, 1995)
   j) ¿Es el Juego de Buena Conducta una estrategia de intervención viable para niños que requieren procedimientos de cambio de comportamiento en el aula? (Tingstrom, 1994)
3. A continuación se presentan diez hipótesis. Discuta las posibilidades de someterlas a prueba. Después, lea dos o tres de los estudios para entender cómo lo hicieron los autores.
   a) Los solicitantes de trabajo que expresan una gran experiencia en tareas no existentes sobreestiman sus habilidades en tareas reales (Anderson, Warner y Spencer, 1984).
   b) En situaciones sociales, los hombres malinterpretan las expresiones amistosas de las mujeres como un signo de interés sexual (Saal, Johnson y Weber, 1989).
   c) A mayor éxito de un equipo, mayor será la atribución que cada miembro haga a su habilidad y suerte personales (Chambers y Abrami, 1991).
   d) El incremento de interés en una tarea aumentará la conformidad (Rind, 1997).
   e) Extractos de la sudoración del hombre pueden afectar el ciclo menstrual de la mujer (Cutler, Preti, Kreiger y Huggins, 1986).
   f) Las personas atractivas físicamente se consideran más inteligentes que las personas no atractivas (Moran y McCullers, 1984).
   g) Uno puede recibir ayuda de un extraño si éste es similar a uno mismo, o si la petición se hace a una cierta distancia (Glick, DeMorest, y Hotze, 1988).
   h) Fumar cigarrillos (nicotina) mejora el desempeño mental (Spilich, June y Remer, 1992).

i) Quienes guardan objetos valiosos en lugares extraños tendrán un mayor recuerdo del sitio que si colocaran estos artículos en lugares comunes (Winograd y Soloway, 1986).

j) Los hombres homosexuales con VIH sintomático presentan significativamente más angustia que aquéllos que desconocen su estatus de VIH (Cochran y Mays, 1994).

4. Los problemas e hipótesis multivariadas (por ahora, más de dos variables dependientes) son ahora comunes en la investigación del comportamiento. Para familiarizar al estudiante con tales problemas, hemos anexado algunos. Trate de imaginar cómo desarrollaría usted la investigación para estudiarlos.

a) ¿Difieren hombres y mujeres en cuanto a sus percepciones acerca de sus genitales, gozo sexual, sexo oral y masturbación? (Reinholtz y Muehlenhard, 1995)

b) ¿Son los fumadores jóvenes más extrovertidos mientras que los fumadores de más edad son más depresivos y aislados? (Stein, Newcomb y Bentler, 1996)

c) ¿Cuánto difiere la apreciación que los maestros tienen de las habilidades sociales de los estudiantes populares y los rechazados? (Frentz, Gresham y Elliot, 1991; Stuart, Gresham y Elliot 1991)

d) ¿Influye el grado de semejanza del consejero y del cliente en cuanto a grupo étnico, género y lenguaje en los resultados del tratamiento para niños de edad escolar? (Hall, Kaplan y Lee, 1994)

e) ¿Existen diferencias en las habilidades cognitivas y funcionales de pacientes con Alzheimer que residen en una unidad de cuidado especial en relación con aquellos que viven en una unidad de cuidados tradicional? (Swanson, Maas y Buckwalter, 1994)

f) ¿Difieren los niños hiperactivos con déficit de atención de los niños no hiperactivos con déficit de atención en cuanto a rendimiento en lectura, ortografía y lenguaje escrito? (Elbert, 1993)

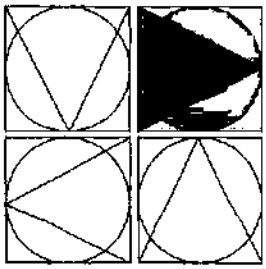
g) ¿La gente ve a las mujeres que prefieren el título de cortesía de señorita como poseedoras de mayores cualidades instrumentales y de menores cualidades de expresividad que las mujeres que prefieren los títulos de cortesía tradicionales? (Dion y Cota, 1991)

h) ¿Aumentará el estilo de liderazgo autoritario la satisfacción de los miembros del grupo? ¿Aumentará la percepción sobre la eficacia del grupo de trabajo su efectividad? (Kumpfer, Turner, Hopkins y Librett, 1993)

i) ¿Cómo influyen el grupo étnico, el género y los antecedentes socioeconómicos en la propensión a la psicosis: aberración perceptiva, ideación mágica y personalidad esquizoide? (Porch, Ross, Hanks y Whitman, 1995)

j) ¿Tendrá la exposición a los estímulos dos efectos, uno cognitivo y otro afectivo, que a su vez afecten la predilección, la familiaridad, certeza y precisión en el reconocimiento? (Zajonc, 1980)

Los últimos dos problemas y estudios resultan muy complejos en tanto que las relaciones establecidas son complejas. Los otros problemas y estudios, aunque complejos, poseen tan sólo un fenómeno presumiblemente afectado por otro, mientras que los últimos dos contienen varios fenómenos que afectan a dos o más fenómenos. El lector no deberá desalentarse si encuentra en ellos algo de dificultad: hacia el final del libro parecerán interesantes y naturales.



CAPÍTULO 3

CONSTRUCTOS, VARIABLES Y DEFINICIONES

- CONCEPTOS Y CONSTRUCTOS
- VARIABLES
- DEFINICIONES CONSTITUTIVAS Y OPERACIONALES DE CONSTRUCTOS Y VARIABLES
- TIPOS DE VARIABLES
  - **Variables independientes y dependientes**
  - **Variables activas y variables atributo**
  - **Variables continuas y categóricas**
- CONSTRUCTOS OBSERVABLES Y VARIABLES LATENTES
- EJEMPLOS DE VARIABLES Y DEFINICIONES OPERACIONALES

Los científicos operan en dos niveles: teoría-hipótesis-constructo y observación. Para ser más exactos, oscilan entre uno y el otro de forma continua. Un psicólogo científico podría decir: "La deprivación temprana produce deficiencia en el aprendizaje." Este enunciado es una hipótesis integrada por dos conceptos, "deprivación temprana" y "deficiencia en el aprendizaje", unidos por una palabra de relación, *produce*. Este enunciado se encuentra en el nivel teoría-hipótesis-constructo. Cuando los científicos formulan enunciados relacionales y utilizan conceptos, o constructos, como los llamaremos, están operando en este nivel.

Los científicos también deben operar en el nivel de observación. Deben reunir datos para probar sus hipótesis. Para hacerlo, es necesario que pasen del nivel de constructo al de observación. No pueden simplemente hacer observaciones de "deprivación temprana" y "deficiencia en el aprendizaje". Es necesario que definan estos constructos de modo que sea posible realizar las observaciones. El problema que se estudia en este capítulo es cómo examinar y aclarar la naturaleza de los conceptos científicos o constructos. Este capítulo también examinará y aclarará la forma en que los científicos del comportamiento pasan del nivel de constructo al de observación, cómo van de uno a otro.

Conceptos y constructos

Los términos "concepto" y "constructo" tienen significados similares, aunque existe una diferencia importante. Un "concepto" expresa una abstracción creada por una generalización a partir de instancias particulares. "Peso" es un concepto que expresa numerosas observaciones de cosas que son "más o menos" y "pesadas o ligeras". "Masa", "energía" y "fuerza" son conceptos usados por científicos de la física. Por supuesto, son mucho más abstractos que conceptos como "peso", "alto" y "longitud".

Un concepto de mayor interés para los lectores es el "aprovechamiento". Es una abstracción que se genera a partir de la observación de ciertos comportamientos de los niños, que se asocian con el dominio del "aprendizaje" en tareas escolares —lectura de palabras, resolución de problemas aritméticos, elaboración de dibujos, etcétera—. Los diversos comportamientos observados se reúnen y expresan en una palabra. "Aprovechamiento", "inteligencia", "agresividad", "conformidad" y "honestidad" son conceptos usados para expresar la variedad del comportamiento humano.

Un *constructo* es un concepto, que tiene el significado agregado de haber sido inventado o adoptado para un propósito científico especial, de forma deliberada y consciente. "Inteligencia" es un concepto, una abstracción de la observación de comportamientos presumiblemente inteligentes y no inteligentes. Sin embargo, como todo constructo científico, "inteligencia" implica tanto más como menos de lo que pueda significar como concepto. Esto quiere decir que los científicos de manera consciente y sistemática la usan en las dos formas: 1) se incorpora en los esquemas teóricos y se relaciona en diversas formas con otros constructos (podemos decir, por ejemplo, que el aprovechamiento escolar es en parte función de la inteligencia y motivación) y 2) "inteligencia" se define y especifica de tal forma que pueda ser observada y medida (podemos hacer observaciones de la inteligencia de los niños al aplicar una prueba de inteligencia, o al solicitar a los maestros que señalen el grado relativo de inteligencia de sus alumnos).

Variables

Los científicos de forma algo vaga, llaman a los constructos o propiedades que estudian, "variables". Algunos ejemplos de variables importantes en sociología, psicología, ciencia política y educación son: género, ingreso, educación, clase social, productividad organizacional, movilidad ocupacional, nivel de aspiración, aptitud verbal, ansiedad, afiliación religiosa, preferencias políticas, desarrollo político (de las naciones), orientación ocupacional, prejuicios raciales y étnicos, conformidad, recuerdo, memoria de reconocimiento y aprovechamiento. Puede decirse que una variable es una propiedad que asume diversos valores. Siendo redundantes, una variable es algo que varía. Aunque esta forma de expresarlo nos aporta una noción intuitiva de lo que son, necesitamos una definición al mismo tiempo más general y precisa.

Una *variable* es un símbolo al que se le asignan valores o números. Por ejemplo, x es una variable: es un *símbolo* al que se le asignan valores numéricos. La variable x puede tomar cualquier conjunto justificable de valores, por ejemplo, puntajes en una prueba de inteligencia o en una escala de actitudes. En el caso de la inteligencia, asignamos a x un conjunto de valores numéricos proporcionados por el procedimiento especificado en una determinada prueba de inteligencia. Este grupo de valores varía de bajo a alto, por ejemplo de 50 a 150.

Una variable, x, sin embargo, puede tener sólo dos valores. Si el género es el constructo bajo estudio, entonces a x se le pueden asignar 1 y 0, donde 1 representa uno de los géne-

ros y 0 el otro. Aún así, es una variable. Otros ejemplos de variables con dos valores son: dentro–fuera, correcto–incorrecto, viejo-joven, ciudadano-no ciudadano, clase media-clase trabajadora, maestro-no maestro, republicano-demócrata, etcétera. Tales variables se llaman *dicotomías*, variables dicotómicas o binarias.

Algunas de las variables usadas en la investigación conductual son verdaderas dicotomías, es decir, se caracterizan por la presencia o ausencia de una propiedad: masculino-femenino, con hogar-indigente, empleado-desempleado. Otras variables son *politomías*. Un buen ejemplo es la preferencia religiosa: protestante, católico, musulmán, judío, budista, otra. Tales dicotomías y politomías se denominan "variables cualitativas". La naturaleza de este calificativo se analizará más adelante. En teoría, muchas variables, sin embargo, pueden asumir valores continuos. Ha sido una práctica común en la investigación del comportamiento convertir las variables continuas en dicotómicas o politómicas. Por ejemplo, la inteligencia, una variable continua, se ha dividido en inteligencia alta, media y bajas, o en alta y baja. Variables como ansiedad, introversión y autoritarismo han recibido un trato similar. Aunque no es posible convertir una variable que de manera natural es dicotómica, como género, en una variable continua, siempre podemos convertir una variable continua en una dicotómica o politómica. Más adelante se verá que tal conversión puede tener un propósito conceptual útil, pero para el análisis de datos constituye una práctica negativa en tanto que descarta información.

Definiciones constitutivas y operacionales de constructos y variables

La diferencia que hicimos antes entre "concepto" y "constructo" conduce de manera natural a otra distinción importante entre los tipos de definiciones de constructos y variables. Se puede definir a las palabras o constructos de dos formas generales: primero, podemos definir una palabra con el uso de otras palabras, que es lo que hace un diccionario. Podemos definir *inteligencia* como "un intelecto operante", "agudeza mental" o "la habilidad para pensar de forma abstracta". Tales definiciones utilizan otros conceptos o expresiones conceptuales en lugar de la expresión o palabra que se define. Segundo, podemos definir una palabra por las acciones o comportamientos que expresa o implica. Definir *inteligencia* de esta forma requiere que especifiquemos qué comportamientos de los niños son "inteligentes" y cuáles son "no inteligentes". Podemos decir que un niño de siete años de edad que lee una historia con éxito es "inteligente", si el niño no puede leer la historia podemos asumir que el chico "no es inteligente". En otras palabras, esta clase de definición puede llamarse *definición observacional* o *conductual*. Se usan definiciones a partir de "otras palabras" y definiciones "observacionales" de manera cotidiana.

En esta discusión hay una imprecisión perturbadora. Aunque los científicos utilizan los tipos de definiciones que acabamos de describir, lo hacen en una forma más precisa. Expresamos este uso al definir y explicar la diferencia que Margenau (1950/1977) plantea para las definiciones constitutivas y operacionales. Una definición *constitutiva* define un constructo usando otros constructos. Por ejemplo, podemos definir *peso* diciendo que es la "pesadez" de los objetos, o *ansiedad* como "temor subjetivo". En ambos casos hemos sustituido un concepto por otro. Algunos de los constructos de una teoría científica pueden definirse de manera constitutiva. Torgerson (1958/1985), a partir de las ideas de Margenau, indica que para ser útiles desde el punto de vista científico, todos los constructos deben poseer significado constitutivo, es decir, poder ser usados en teorías.

Una definición *operacional* asigna significado a un constructo o variable al especificar las actividades u "operaciones" necesarias para medirlo y evaluar la medición. De manera

alternativa, una definición operacional constituye una especificación de las actividades del investigador para medir una variable o para manipularla. Implica algo así como un manual de instrucciones para el investigador. En efecto, dice, "haga tal y cual, de la forma tal y tal". En síntesis, define o aporta significado a una variable al delinear paso a paso lo que el investigador debe hacer para medirla y para evaluar dicha medición.

Michel (1990) presenta una excelente revisión histórica de cómo las definiciones operacionales se hicieron populares en las ciencias sociales y del comportamiento. Michel cita a P. W. Bridgeman, premio Nobel, como creador de la definición operacional en 1927. Bridgeman, como lo relata Michel (1990, p. 15), indica: "en general, con cualquier concepto dado queremos significar tan sólo un conjunto de operaciones; *el concepto es sinónimo del correspondiente conjunto de operaciones*". Cada operación diferente definiría un concepto distinto.

Un ejemplo bien conocido, aunque extremo, de una definición operacional es: inteligencia (ansiedad, aprovechamiento, etcétera) es la puntuación en una prueba *X* de inteligencia, o inteligencia es lo que la prueba *X* de inteligencia mide. Las puntuaciones altas indican un mayor nivel de inteligencia que las bajas. Esta definición indica qué hacer para medir la inteligencia y no precisa qué tan bien es medida la inteligencia por el instrumento especificado. (Se presume que se indagó sobre la adecuación de esta prueba antes de que el investigador la usara.) En este tipo de uso, una definición operacional equivale a una ecuación en la que decimos: "si la inteligencia es igual a la puntuación en la prueba *X* de inteligencia, las puntuaciones altas indican un mayor grado de inteligencia que los bajos". Parecería también que estamos diciendo "el significado de inteligencia (en este estudio) se expresa por el puntaje en la prueba *X* de inteligencia".

Existen, en general, dos clases de definiciones operacionales: 1) las *medidas* y 2) las *experimentales*. La definición de arriba está más estrechamente ligada con las definiciones medidas que con las experimentales. Una definición operacional *medida* describe cómo será medida una variable. Por ejemplo, aprovechamiento puede definirse por una prueba estandarizada de aprovechamiento, por un examen desarrollado por el maestro, o por las calificaciones. Doctor, Cutris e Isaacs (1994), al estudiar el efecto de la consejería para el estrés en oficiales de policía, definieron de manera operacional morbilidad psiquiátrica como las puntuaciones en el Cuestionario General de Salud y el número de días que habían tomado por incapacidad. Altas puntuaciones y un gran número de días indicaban niveles elevados de morbilidad. Little, Sterling y Tingstrom (1996) estudiaron los efectos de la raza y origen geográfico en la atribución. La atribución se definió operacionalmente como la puntuación en el cuestionario de estilo atribucional. Un estudio puede incluir la variable *consideración* que puede definirse operacionalmente a través de una lista de comportamientos de niños que se presume son comportamientos considerados. Después, se puede pedir a los maestros que evalúen a los alumnos en una escala de cinco puntos. Tales comportamientos pueden ejemplificarse como instancias en que un niño le dice a otro "lo siento" o "discúlpame", o cuando un chico le presta un juguete a otro o cuando se lo pide (sin la amenaza de agresión), o cuando un pequeño ayuda a otro con una tarea escolar. También puede definirse por la suma de comportamientos considerados: a mayor cantidad, mayor nivel de consideración.

Una definición operacional *experimental* señala los detalles (operaciones) de la manipulación de una variable por parte del investigador. El reforzamiento puede definirse operacionalmente al precisar los detalles de cómo los sujetos serán reforzados (premiados) y no reforzados (no premiados) por comportamientos específicos. Hom, Berger, Duncan, Miller y Belvin (1994) definieron operacionalmente el reforzamiento en forma experimental. En este estudio, los niños fueron asignados a uno de cuatro grupos. Dos de los grupos estuvieron sujetos a una condición de reforzamiento con base en cooperación,

mientras que los otros dos trabajaron en un esquema en que se reforzaba la postura individualista. Bahrick (1984) definió la memoria a largo plazo en términos de al menos dos procesos en lo referente a la retención de información de tipo académico. Un proceso llamado "almacenaje permanente" (*permastore*), elige de manera selectiva alguna información para ser almacenada de forma permanente y resulta muy resistente al decaimiento (olvido). El otro proceso al parecer elige información menos significativa por lo que es menos resistente al olvido. Esta definición contiene implicaciones claras para la manipulación experimental. Strack, Martin y Stepper (1988) definieron operacionalmente la sonrisa como la activación de los músculos asociados con la sonrisa humana. Lo lograron al solicitar que una persona sostuviera una pluma en su boca de cierta manera. Se trata de un procedimiento no intrusivo en tanto que no se les pidió a los participantes que posaran con una sonrisa. Se presentarán otros ejemplos de ambos tipos de definiciones más adelante.

Los investigadores científicos eventualmente enfrentan la necesidad de medir las variables de las relaciones que estudian. Algunas mediciones son fáciles, otras difíciles. Medir el género o la clase social es fácil; pero evaluar creatividad, conservadurismo o efectividad organizacional resulta difícil. La importancia de las definiciones operacionales no puede dejar de enfatizarse. Ellas son ingredientes indispensables de la investigación científica porque permiten al investigador medir variables y porque representan puentes entre el nivel de la teoría-hipótesis-constructo y el nivel de observación. No hay investigación científica sin observaciones, y éstas no son posibles sin instrucciones claras y específicas de qué y cómo observar. Las definiciones operacionales son tales instrucciones.

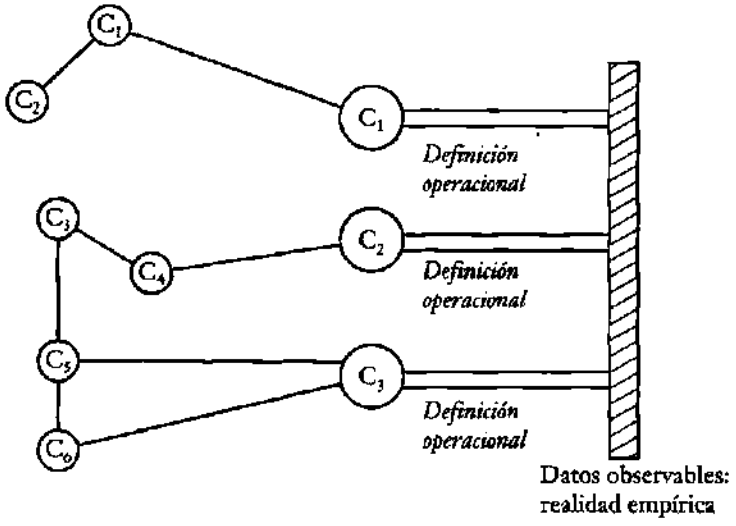
Aunque indispensables, las definiciones operacionales sólo aportan significados limitados de los constructos. Ninguna definición operacional puede expresar toda la riqueza y los diversos aspectos de algunas variables, como sucede con el prejuicio humano. Esto implica que las variables medidas por los científicos siempre tienen un significado limitado y específico. La "creatividad" que estudian los psicólogos no se refiere necesariamente a la "creatividad" de los artistas, aunque, por supuesto, tengan elementos comunes. Una persona que piensa en una solución creativa para un problema matemático puede mostrar una escasa creatividad como poeta (Barron y Harrington, 1981). Algunos psicólogos han definido operacionalmente a la creatividad como el desempeño en la prueba de Torrance de Pensamiento Creativo (Torrance, 1982). Los niños que obtienen una puntuación alta en esta prueba, tienen mayor probabilidad de exhibir logros creativos en la edad adulta.

Algunos científicos afirman que estos limitados significados operacionales son los únicos significados que "significan" algo, que todas las demás definiciones son disparates metafísicos. Señalan que las discusiones sobre ansiedad constituyen tonterías de tipo metafísico, a menos que se cuente con y se usen, definiciones operacionales adecuadas. Éste es un punto de vista extremo, aunque posee algunos aspectos saludables. Insistir en que cada término que usemos en el discurso científico sea definido operacionalmente sería demasiado reduccionista y restrictivo y, como se verá, científicamente erróneo. Northrop (1947/1983, p. 130) señala, por ejemplo: "La importancia de las definiciones operacionales estriba en que hacen posible la verificación y enriquecen el significado. Sin embargo, no agotan el significado científico". Margenau (1950/1977, p. 232) señala el mismo punto en su extensa discusión sobre los constructos científicos.

A pesar de los riesgos del operacionalismo extremo, parece seguro decir que ha sido y es una influencia saludable. Como señala Skinner (1945, p. 274):

> La actitud operacional, a pesar de sus inconvenientes, es positiva en cualquier ciencia, pero en especial en psicología, por la presencia de un vasto vocabulario de origen antiguo y no científico.

FIGURA 3.1



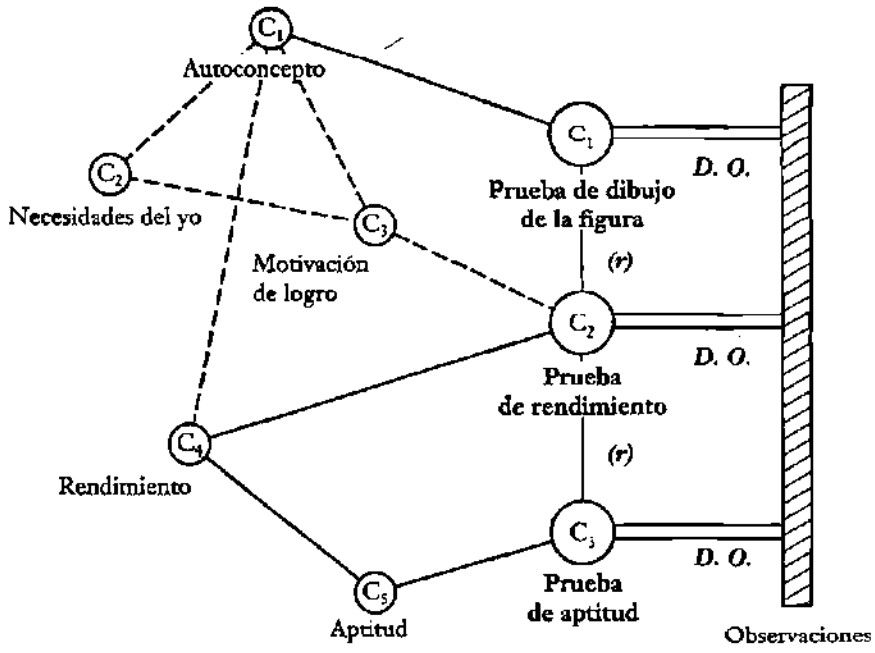
Cuando se consideran los términos usados en educación, es claro que esta disciplina también posee un vasto vocabulario de origen antiguo y no científico. Consideremos éstos: el niño integral, el enriquecimiento horizontal y vertical, la satisfacción de las necesidades del alumno, el tronco común, el ajuste emocional y el enriquecimiento curricular. Esto también se aplica al campo de la atención geriátrica. Aquí la enfermera especializada maneja términos como el proceso de envejecimiento, la autoimagen, el mantenimiento de la atención y la negligencia unilateral (Eliopoulos, 1993; Smeltzer y Bare, 1992).

Para aclarar las definiciones constitutivas y operacionales (así como la teoría) veamos la figura 3.1, que ha sido adaptada de Margenau (1950/1977) y de Torgerson (1958/1985). El diagrama intenta ilustrar una teoría bien desarrollada. Las líneas sencillas representan conexiones teóricas o relaciones entre constructos. Estos constructos, etiquetados con letras minúsculas, se definen de manera constitutiva; esto es, c_4 está definida de alguna forma por c_3 o viceversa. Las líneas dobles representan definiciones operacionales. Los constructos están ligados de forma directa a datos observables, y son vinculaciones indispensables con la realidad empírica. Sin embargo, no todos los constructos en una teoría científica se definen operacionalmente. De hecho, la teoría que tiene todos sus constructos así definidos es algo tenue.

Construyamos una "pequeña teoría" del bajo rendimiento para ilustrar estos conceptos. Supongamos que un investigador cree que el bajo rendimiento es en parte una función del autoconcepto de los alumnos. Piensa que los estudiantes que se perciben como inadecuados y que tienen percepciones negativas de sí mismos, también tienden a rendir menos que lo que su capacidad potencial y aptitudes indican. De aquí se sigue que las necesidades del yo (que no definiremos aquí) y la motivación de logro (que llamaremos necesidad de logro) están ligadas al bajo rendimiento. Como es natural, el investigador está consciente de la relación entre aptitud e inteligencia y logro en general. Un diagrama que ilustra esta "teoría" puede ser como el de la figura 3.2.

El investigador no cuenta con una medida *directa* de autoconcepto, pero supone que puede inferirlo a partir de una prueba de dibujo de la figura. Entonces define operacional-

FIGURA 3.2



mente el autoconcepto como ciertas respuesta a esa prueba. Éste es probablemente el método más común para medir constructos psicológicos (y educativos). La línea gruesa entre c_1 y C_1 indica la naturaleza relativamente directa de la relación que se asume entre el autoconcepto y la prueba. (La línea doble entre C_1 y el nivel de observación indica una definición operacional, como en la figura 3.1.)

De forma parecida, el constructo de rendimiento (c_4) se define operacionalmente como la discrepancia entre la medición realizada del rendimiento (C_2) y la de la aptitud (c_5). En este modelo el investigador no mide directamente la motivación de logro, ni tampoco posee una definición operacional de ella. En otro estudio se puede hipotetizar específicamente una relación entre rendimiento y motivación de logro, en cuyo caso se tratará de definir motivación para el logro en forma operacional.

Una línea continua sencilla entre conceptos, por ejemplo la que está entre el constructo rendimiento (c_4) y prueba de aprovechamiento (C_2), indica una relación relativamente bien establecida entre el rendimiento postulado y lo que miden las pruebas estandarizadas de rendimiento. La líneas continuas sencillas entre C_1 y C_2 y aquellas entre C_2 y C_3 indican relaciones obtenidas entre las puntuaciones de las pruebas de estas mediciones. (Las líneas entre C_1 y C_2, y entre C_2 y C_3 están marcadas como (r) para significar "relación" o "coeficiente de correlación".)

Las líneas discontinuas representan relaciones postuladas entre constructos que no están relativamente bien establecidas. Un buen ejemplo de esto es la relación postulada entre el autoconcepto y la motivación de logro. Uno de los propósitos de la ciencia es convertir estas líneas punteadas en continuas al cerrar la brecha entre definición-opera-

cional-medición. En este caso, es concebible que tanto el autoconcepto como la motivación de logro puedan ser definidas operacionalmente y medidas directamente.

En esencia, ésta es la forma en que el científico del comportamiento opera. Este especialista se desplaza de manera continua entre el nivel del constructo y el de la observación y lo logra al definir operacionalmente las variables de la teoría que pueden serlo. Luego, se estiman las relaciones entre la definición operacional y las variables medidas. De estas relaciones estimadas, el científico abstrae inferencias acerca de las relaciones entre los constructos. En el ejemplo anterior, el científico del comportamiento calcula la relación entre C_1 (la prueba de dibujo de la figura) y C_2 (prueba de rendimiento). Si la relación se establece en este nivel observacional, el científico infiere que existe una relación entre c_1 (autoconcepto) y c_4 (rendimiento).

Tipos de variables

Variables independientes y dependientes

Dejemos atrás los fundamentos de las definiciones para regresar a las variables. Es posible clasificar las variables de diversas formas. En este libro, tres tipos de variables son muy importantes y serán enfatizadas: 1) variables independientes y dependientes, 2) variables activas y atributo, y 3) variables continuas y categóricas.

La forma más útil de categorizar las variables es como independientes o dependientes. Esta taxonomía resulta muy útil debido a su aplicabilidad general, su simplicidad y su importancia especial tanto en la conceptualización como en el diseño de la investigación, así como en la comunicación de los resultados de ésta. Una *variable independiente* es la causa *supuesta* de la *variable dependiente*, el efecto *supuesto*. La variable independiente es el antecedente; la dependiente es el consecuente. Dado que uno de los objetivos de la ciencia es descubrir relaciones entre diferentes fenómenos, la búsqueda de las relaciones entre variables independientes y dependientes lo logra. Se asume que la variable independiente influye en la dependiente. En algunos estudios, la variable independiente "causa" cambios en la variable dependiente. Cuando decimos: si *A*, entonces *B*, tenemos una conjunción condicional de una variable independiente (*A*) y una variable dependiente (*B*).

Los términos "variable independiente" y "variable dependiente" proceden de las matemáticas, donde X es la variable independiente y Y es la dependiente. Ésta es probablemente la mejor forma de pensar en las variables independientes y dependientes porque no hay necesidad de utilizar la discutida palabra *causa* y las palabras afines a ella, y dado que el uso de tales símbolos se aplica a la mayoría de las situaciones de investigación. No hay restricción teórica alguna en cuanto a la cantidad de X y Y. Cuando más adelante consideremos el pensamiento y análisis multivariado, trataremos con diversas variables dependientes e independientes.

En los experimentos, el investigador manipula la variable independiente. Los cambios en los valores o niveles de la variable independiente generan cambios en la variable dependiente. Cuando investigadores del campo educativo estudiaron los efectos de diferentes métodos de enseñanza en el desempeño en una prueba de matemáticas, ellos variaron los métodos de enseñanza. En una condición pudieron tener "sólo exposición frontal", en la otra pudo haber "exposición frontal y video". El método de enseñanza es la variable independiente. La variable resultado, la puntuación en la prueba de matemáticas, es la variable dependiente.

La asignación de participantes a diferentes grupos con base en la existencia de alguna característica es un ejemplo de cuando el investigador no puede manipular la variable

independiente. En esta situación, los valores de la variable independiente son preexistentes. El participante tiene la característica o no la tiene. En este caso, no hay posibilidad de una manipulación experimental, pero se considera que "lógicamente" la variable tiene algún efecto en la variable dependiente. Las variables de características del sujeto constituyen la mayor parte de este tipo de variables independientes. Una de las variables independientes más comunes de este tipo es el género (femenino y masculino). Así, si un investigador desea determinar si hombres y mujeres difieren en las destrezas matemáticas, se aplicaría una prueba matemática a representantes de ambos grupos, y se compararían las puntuaciones de la prueba. La prueba de matemáticas sería la variable dependiente. Una regla general es que cuando el investigador manipula una variable o asigna participantes a los grupos según alguna característica, esa variable es la independiente. La tabla 3.1 muestra una comparación entre los dos tipos de variables independientes y su relación con la variable dependiente. La variable independiente debe tener al menos dos niveles o valores. Observe en la tabla 3.1 que ambas situaciones presentan dos niveles para la variable independiente.

La variable dependiente es, por supuesto, *hacia* la que se hace la predicción, mientras que la independiente es aquella *a partir de* la cual se predice. La variable dependiente, Y, es el efecto supuesto, que varía de manera concomitante a los cambios o variaciones en la variable independiente, X; es la variable que se observa para detectar variaciones como un resultado supuesto de la variación en la variable independiente. La variable dependiente es el resultado medido que el investigador usa para determinar si los cambios en la variable independiente tuvieron un efecto. Al predecir Y a partir de X, podemos tomar cualquier valor de X que deseemos, mientras que el valor de Y que predecimos *es* "dependiente" del valor de X que hemos elegido. En general, la variable dependiente es la condición que tratamos de explicar. Por ejemplo, la variable dependiente más común en educación, es "aprovechamiento" o "aprendizaje". Deseamos explicar o dar cuenta del aprovechamiento. Para ello tenemos un gran número de posibles X o variables independientes de dónde elegir.

Cuando se estudia la relación entre inteligencia y aprovechamiento escolar, la inteligencia es la variable independiente y el aprovechamiento es la dependiente. (¿Se podría concebir a la inversa?) Otras variables independientes que pueden estudiarse con relación a la variable dependiente aprovechamiento son: clase social, métodos de enseñanza, tipos de personalidad, tipos de motivación (recompensa y castigo), actitudes hacia la escuela y ambiente en el salón de clases, entre otros. Cuando se estudian los supuestos determinantes de la delincuencia, aquellos tales como condiciones de pobreza, hogares desintegrados, falta de amor de los padres y aspectos similares, constituyen las variables independientes y,

TABLA 3.1 *Relación de variables independientes manipuladas y no manipuladas con la variable dependiente*

	Niveles de la variable independiente			
	Método de enseñanza		**Género**	
	Sólo exposición frontal Puntuaciones en la prueba de matemáticas	Exposición frontal y video Puntuaciones en la prueba de matemáticas	Femenino Puntuaciones en la prueba de matemáticas	Masculino Puntuaciones en la prueba de matemáticas
	Variables dependientes		Variables dependientes	

como es natural, la delincuencia (o mejor aún, el comportamiento delictivo) es la variable dependiente. En la hipótesis frustración-agresión, frustración es la variable independiente, y agresión, la dependiente. En ocasiones, un fenómeno se estudia por sí mismo y ya sea la variable dependiente o la independiente están implícitas. Es el caso en que se estudian los comportamientos y características del maestro. La variable dependiente implícita común es el aprovechamiento o el comportamiento del niño, aunque el comportamiento del maestro puede, por supuesto, constituir una variable dependiente. Consideremos un ejemplo en el campo de la atención médica. Cuando se comparan medidas cognitivas y funcionales de pacientes con Alzheimer entre instituciones de internamiento tradicional y unidades de cuidado especial, la variable independiente es el lugar de atención. Las variables dependientes son las medidas cognitivas y funcionales (Swanson, Maas y Buckwalter, 1994).

La relación entre una variable independiente y una dependiente se puede entender mejor si trazamos dos ejes perpendiculares uno del otro. Uno representa a la variable independiente; el otro, a la dependiente. (Cuando dos ejes forman ángulos rectos entre sí se denominan ejes *ortogonales*.) De acuerdo a la tradición matemática, x, la variable independiente, es el eje horizontal y y, la dependiente, representa el eje vertical (x se denomina la abscisa y y la ordenada). Los valores para x se grafican en el eje de las x, y los valores de y en el eje de las y.

Una forma común y útil de "ver" e interpretar una relación es graficar un par de valores de xy, usando los ejes x y y como marco de referencia. En un estudio de desarrollo infantil, supongamos que tenemos dos grupos de medidas. Las medidas x de edad cronológica y las medidas y representan edad lectora. *La edad lectora* se denomina también edad de crecimiento. Las mediciones en serie de diferentes áreas del crecimiento de los individuos —por ejemplo estatura, peso o inteligencia— se expresan como la edad cronológica promedio en la que aparecen en la población estándar.

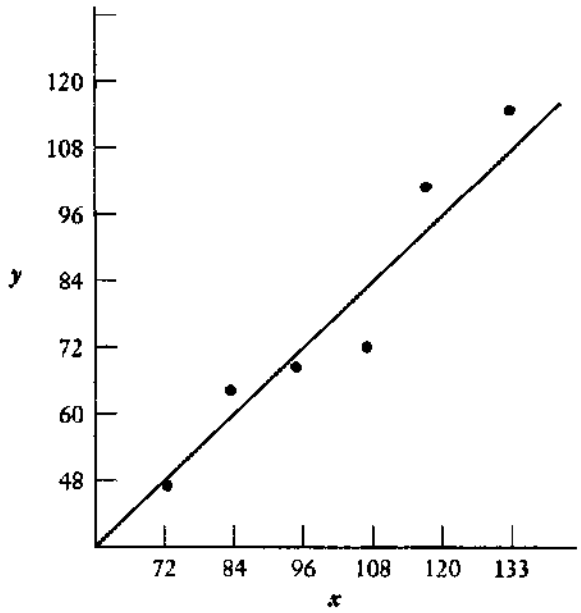
x: edad cronológica (en meses)	y: edad lectora (en meses)
72	48
84	62
96	69
108	71
120	100
132	112

Estas medidas se grafican en la figura 3.3.

La relación entre edad cronológica (EC) y edad lectora (EL), ahora puede "verse" y aproximarse de forma burda. Observe que hay una tendencia pronunciada (como se podría esperar), para que una mayor EC se asocie con una mayor EL, una EC media con una EL media, y una EC menor, con una EL menor. En otras palabras, la relación entre las variables independiente y dependiente, en este caso EC y EL, puede observarse en una gráfica como la que aparece en la figura 3.3. Se ha trazado una recta para "mostrar" esta relación: constituye un promedio aproximado de todos los puntos de la gráfica. Observe que si uno conoce las medidas de la variable independiente y una relación como la que se muestra en la figura 3.3, uno puede predecir, con considerable precisión, las medidas de la variable dependiente. Gráficas como ésta pueden usarse, desde luego, con cualesquier grupo de medidas para variables dependientes e independientes.

El estudiante debe estar atento a la posibilidad de que en un estudio una variable sea independiente, mientras en otro sea dependiente, e inclusive ambas en un mismo estudio.

FIGURA 3.3



Un ejemplo es la satisfacción laboral. La mayoría de los estudios sobre satisfacción laboral la utilizan como variable dependiente. Day y Schoenrade (1997) muestran el efecto de la orientación sexual en las actitudes laborales. Una de estas actitudes laborales es la satisfacción laboral. De la misma forma, Lekewise, Hodson (1989) estudia las diferencias de género en la satisfacción laboral. Scott, Moore y Miceli (1997) encuentran a la satisfacción laboral ligada a los patrones de comportamiento de los adictos al trabajo. Hay estudios en donde la satisfacción laboral es usada como una variable independiente: Meiksins y Watson (1989) muestran cuánto influye la satisfacción laboral en la autonomía profesional de los ingenieros. Estudios de Somers (1996); Francis-Felsen, Coward, Hogan y Duncan (1996); y Hutchinson y Turner (1988) evaluaron el efecto de la satisfacción laboral en la rotación del personal de enfermería.

Otro ejemplo es la ansiedad, que se ha estudiado como una variable independiente que afecta a la variable dependiente aprovechamiento. Oldani (1997) encontró que la ansiedad de la madre durante el embarazo influye en el aprovechamiento de los hijos (medido como éxito en el área musical). Capaldi, Crosby y Stoolmiller (1996) emplearon los niveles de ansiedad de varones adolescentes para predecir el momento de su primer encuentro sexual. Onwuegbuzie y Seaman (1995) estudiaron los efectos de la ansiedad ante los exámenes en la realización de la prueba, en un curso de estadística. La ansiedad también puede concebirse y usarse como una variable dependiente: por ejemplo, puede utilizarse para estudiar la diferencia entre tipos de cultura, nivel socioeconómico y género (véase Guida y Ludlow, 1989; Murphy, Olivier, Monson y Sobol, 1991). En otras palabras, la clasificación de la variable independiente y la dependiente es en realidad una taxonomía de los usos de la variable más que una distinción entre diferentes tipos de variables.

Variables activas y variables atributo

Una clasificación que nos será útil en nuestro estudio posterior del diseño de la investigación se basa en la distinción entre variables experimentales y medidas. Cuando se planea y ejecuta la investigación es importante distinguir entre estos dos tipos de variables. Las variables manipuladas se llamarán variables *activas*, mientras que las variables medidas se denominarán variables *atributo*. Por ejemplo, Colwell, Foreman y Trotter (1993) compararon dos métodos de tratamiento de las úlceras de presión de los pacientes encamados. Las variables dependientes fueron eficacia y efectividad costo. Los dos métodos de tratamiento fueron una compresa de gasa humedecida y una compresa con una cubierta de hidrocoloide. Los investigadores controlaron quién recibía qué tipo de tratamiento. Como tal, el tratamiento o variable independiente fue una variable activa o manipulada.

Así, cualquier variable manipulada, constituye una variable activa. "Manipulación" significa, en esencia, hacer cosas diferentes a distintos grupos de sujetos, como se verá con claridad en un capítulo posterior al discutir a profundidad las diferencias entre investigación experimental y no experimental. Se dice que existe manipulación cuando un investigador hace algo a un grupo (por ejemplo, reforzar positivamente cierta clase de comportamiento) y hace algo distinto con el otro grupo, o tiene a dos grupos siguiendo diferentes instrucciones. Cuando uno usa diferentes métodos de enseñanza o premia a los sujetos de un grupo y castiga a los de otro, o crea ansiedad a través de instrucciones que generan preocupación, uno está *activamente* manipulando las variables: métodos, reforzamiento y ansiedad.

Otra clasificación relacionada, usada principalmente por los psicólogos es la de las variables *estímulo* y *respuesta*. Una *variable estímulo* es cualquier condición o manipulación del ambiente realizada por el experimentador, que evoque una respuesta en un organismo. Una *variable de respuesta* es cualquier clase de comportamiento del organismo. El supuesto es que para cualquier clase de comportamiento siempre hay un estímulo. Por lo tanto, el comportamiento del organismo es una respuesta. Esta clasificación se refleja en la bien conocida ecuación: $R = f(O, E)$, que se lee: "las respuestas son una función del organismo y los estímulos", o "las variables de respuesta son una función de las variables organísmicas y de las variables estímulo".

Las variables que no pueden ser manipuladas son *las atributo o características del sujeto*. Es imposible, o al menos muy difícil, manipular muchas variables. Las variables consistentes en características humanas como inteligencia, aptitud, género, estatus socioeconómico, conservadurismo, dependencia de campo, necesidad de logro y actitudes son variables atributo. Los sujetos llegan a nuestro estudio con estas variables (atributos) ya presentes o preexistentes. El entorno temprano, la herencia, y otras circunstancias han hecho de los individuos lo que son. Se les llama también variables *organísmicas*. Cualquier propiedad, característica o atributo de un individuo constituye una variable organísmica, digamos que forma parte del organismo. En otras palabras, las variables organísmicas son aquellas características que los individuos poseen en diversos grados cuando ingresan a la situación de investigación. El término *diferencias individuales* implica variables organísmicas. Una de las más comunes variables atributo en las ciencias sociales y del comportamiento es el género: femenino-masculino. Los estudios diseñados para comparar diferencias de género involucran una variable atributo. Tomemos, por ejemplo, el estudio de Weerth y Kalma (1993). Estos investigadores compararon a hombres y mujeres en su respuesta a la infidelidad del cónyuge o pareja. La variable atributo aquí es género, que *no* es una variable manipulada. Hay estudios donde las puntuaciones de una o varias pruebas se usaron para dividir a un conjunto de personas en dos o más grupos. En este caso, las diferencias del grupo se reflejan como una variable atributo, como lo ilustra el estudio de Hart, Forth y

Hare (1990) quienes aplicaron una prueba de psicopatología a varones recluidos en una prisión. Con base en sus puntuaciones, los internos fueron asignados a uno de tres grupos: bajo, medio y alto. Después se comparó su puntuación en una batería de pruebas neuropsicológicas. El nivel de psicopatología preexiste y el investigador no lo manipula. Si un interno puntuaba alto, era asignado al grupo alto. Así, la psicopatología es una variable atributo en este estudio. Hay algunos estudios donde la variable independiente podría haber sido manipulada; sin embargo, por razones logísticas o legales no lo fue. Un ejemplo es el estudio de Swanson, Maas y Buckwalter (1994) quienes compararon diferentes formas de atención y su efecto en las medidas cognitivas y funcionales de los pacientes con Alzheimer. La variable atributo fue el tipo de atención. No se permitió a los investigadores asignar a sus pacientes a dos diferentes instituciones de atención (tradicional *vs.* unidad de cuidado especial). Los investigadores se vieron forzados a estudiar a los sujetos una vez que habían sido asignados al centro correspondiente. Por ello puede considerarse que la variable independiente es una variable no manipulada. Los investigadores heredaron grupos intactos.

La palabra ***atributo*** es lo suficientemente precisa cuando se usa con objetos o referentes inanimados. Sin embargo, las organizaciones, instituciones, grupos, poblaciones, casas y áreas geográficas también poseen atributos —***atributos activos***—. Las organizaciones son productivas de manera variable; las instituciones pasan de moda; los grupos difieren en su cohesión; las áreas geográficas varían ampliamente en sus recursos.

Esta distinción de atributo activo es general, flexible y útil. Veremos que algunas variables, por su propia naturaleza, son siempre atributos, mientras que otras variables que son atributos pueden también ser activas. Esta última característica hace posible investigar las "mismas" relaciones de diferentes formas. Y usando de nuevo el ejemplo de la variable ansiedad, podemos medirla: como es evidente en este caso, es una variable atributiva. Sin embargo, también puede ser manipulada al inducir diferentes grados de ansiedad: por ejemplo, si les decimos a los sujetos de un grupo experimental que la tarea que van a realizar va a ser muy difícil, que su inteligencia será evaluada y que su futuro depende de la puntuación que obtengan. A los sujetos del otro grupo experimental les decimos que lo hagan lo mejor posible, pero relajados, y que el resultado no es importante y que no va a influir en su futuro. En realidad no podemos asumir que la ansiedad medida (atributo) y la ansiedad manipulada (activa) sean la misma. Podemos suponer que ambas son "ansiedad" en un sentido amplio, pero ciertamente no son iguales.

Variables continuas y categóricas

Ya se ha hecho una distinción entre variables continuas y categóricas en especial útil para la planeación de la investigación y el análisis de datos. Sin embargo, su importancia justifica una consideración más amplia.

Una variable ***continua*** es capaz de asumir un conjunto ordenado de valores dentro de cierto rango. Esta definición significa, primero, que el valor de una variable continua refleja al menos un orden categórico, que un mayor valor de la variable implica más de la propiedad en cuestión que un menor valor. Los valores producidos por una escala para medir dependencia, por ejemplo, expresan diferentes cantidades de dependencia, desde la alta pasando por la media hasta la baja. Segundo, las medidas continuas en uso están contenidas en un rango, y cada individuo obtiene una "puntuación" dentro del mismo. Una escala para medir dependencia puede tener un rango de uno a siete. La mayoría de las escalas usadas en ciencias del comportamiento también tienen una tercera característica: hay un conjunto teóricamente infinito de valores en el rango. (Las escalas de rangos

ordenados son algo diferentes; se analizarán más adelante en el libro.) Esto quiere decir que una puntuación de un individuo en particular puede ser de 4.72 más que sólo de 4 o 5.

Las variables *categóricas*, como las llamaremos, pertenecen a una clase de mediciones llamadas nominales (se explicarán en el capítulo 25). En una medición nominal hay dos o más subconjuntos del grupo de objetos que se mide. Se categoriza a los individuos en razón de la posesión de las características que definen cualquier subgrupo. "Categorizar" significa asignar un objeto a una subclase (o subconjunto) de una clase (o conjunto) con base en que el objeto posea o no la característica que define al subconjunto. El individuo que está en proceso de ser categorizado posee o no la propiedad definitoria; esto es un asunto de todo o nada. Los ejemplos más simples son las variables categóricas dicotómicas: femenino-masculino, republicano-demócrata, correcto-incorrecto. Las variables politómicas, que son aquellas con más de dos subconjuntos, son bastante comunes, en especial en sociología y economía: preferencia religiosa, nivel educativo, nacionalidad y elección laboral, entre otras.

Las variables categóricas y la medición nominal tienen requisitos simples: se considera iguales a todos los miembros de un subconjunto y todos tienen asignado el mismo nombre (nominal) y el mismo número. Si la variable es preferencia religiosa, por ejemplo, todos los protestantes son iguales, todos los católicos son iguales y todos los "otros" son iguales. Si un individuo es católico (definido operacionalmente de una manera apropiada), la persona es asignada a la categoría "católica" y también se le asigna un "1" en esa categoría. Es decir, se contabiliza a ese individuo como "católico". Las variables categóricas son "democráticas". No hay un orden en cuanto a rango, ni mayor que o menor que, entre las categorías y se asigna un mismo valor a todos los miembros de una categoría.

La expresión "variables cualitativas" algunas veces se ha aplicado a las variables categóricas, en particular a las dicotómicas, probablemente en contraste con las "variables cuantitativas" (nuestras variables continuas). Este uso refleja un concepto distorsionado de lo que son las variables, ya que siempre son cuantificables; si no lo fueran, no serían variables. Si x tiene solamente dos subconjuntos y puede asumir sólo dos valores (1 y 0), éstos siguen siendo valores, y la variable varía. Si x es una variable politómica, como la afiliación política, cuantificamos nuevamente al asignar valores enteros a los individuos. Si un individuo dice ser demócrata, se ubica a esa persona en el subgrupo demócrata y se le asigna un 1. Todos los individuos en el subconjunto demócrata tendrán un valor de 1. Es en extremo importante comprender esto porque, por principio, son las bases para cuantificar muchas variables, incluso tratamientos experimentales, para llevar a cabo análisis complejos. En el análisis de regresión múltiple, como veremos más adelante, todas las variables —continuas y categóricas— son ingresadas como variables al análisis. En el ejemplo de género que se mencionó arriba, 1 era asignado a un género y 0 al otro. Diseñamos una columna de unos y ceros de la misma forma en que estableceríamos una columna de puntuaciones de dependencia. La columna de unos y ceros representa la cuantificación de la variable género. Aquí no hay misterio. Estas variables han sido llamadas variables *prototipo* (conocidas como dummy en inglés). Como son muy útiles y poderosas e incluso indispensables en el análisis de datos de la investigación moderna, requieren ser entendidas con claridad. Una explicación más a fondo puede encontrarse en Kerlinger y Pedhazur (1973) y en el capítulo 34 de este libro. El método se aplica con facilidad a las politomías. Una *politomía* es una división de los miembros de un grupo en tres o más subdivisiones.

Constructos observables y variables latentes

En gran parte de la discusión previa de este capítulo se ha implicado —pues no se ha declarado de forma explícita— que existe una diferencia fundamental entre constructos y

variables observadas. Más aún, podemos afirmar que los constructos no son observables; y que las variables, cuando se definen operacionalmente, son observables. Esta distinción es importante en tanto que si no estamos plenamente conscientes del nivel de discurso en que nos encontramos al hablar acerca de variables, es difícil ser claros sobre lo que hacemos.

Una expresión fructífera e importante que se encontrará y usará extensamente en este libro es "variable latente". Una variable latente es una "entidad" no observada, que se presume subyace a las variables observadas. El ejemplo mejor conocido de una variable latente importante es "inteligencia". Podemos decir que tres pruebas de habilidad —verbal, numérica y espacial— están relacionadas de manera positiva y sustancial. Esto significa, en general que las personas con puntuación alta en una, tienden a tener altas puntuaciones en las otras; de forma similar quienes obtienen bajas puntuaciones en una, tenderán a presentar bajas en las otras. Creemos que hay algo común a las tres pruebas o variables observadas, y lo llamamos "inteligencia", que es una variable latente.

Hemos encontrado muchos ejemplos de variables latentes en las páginas previas: aprovechamiento, creatividad, clase social, satisfacción laboral, preferencia religiosa, etcétera. De hecho, siempre que mencionamos los nombres de fenómenos en que varían las personas o los objetos, hablamos de variables latentes. En el campo de la ciencia, nuestro interés real está más en las relaciones entre variables latentes que entre variables observadas, ya que buscamos explicar fenómenos y sus relaciones. Cuando enunciamos una teoría, enunciamos en parte relaciones sistemáticas entre variables latentes. No estamos demasiado interesados en la relación entre el comportamiento observado de frustración y la conducta observada de tipo agresivo, por ejemplo, aunque debemos trabajar con ellos en el nivel empírico. En realidad, estamos interesados en la relación entre la variable latente frustración y la variable latente agresión.

Debemos ser precavidos, sin embargo, cuando tratamos con variables no observables. Los científicos que usan términos como "hostilidad", "ansiedad" y "aprendizaje", están conscientes de que hablan a cerca de constructos inventados. La "realidad" de estos constructos se infiere a partir del comportamiento. Si desean estudiar diferentes tipos de motivación, deben saber que "motivación" es una variable latente, un constructo inventado para dar cuenta de un comportamiento presumiblemente "motivado". Es necesario que sepan que esta "realidad" sólo está postulada. Sólo pueden juzgar si los jóvenes están motivados o no al observar sus comportamientos. Aun así para estudiar la motivación, deben medirla o manipularla. Pero no pueden medirla de forma directa por que, en pocas palabras, es una variable que está "en la cabeza", una entidad no observable, una variable latente. Se inventó el constructo "por algo" que *se presume que está* dentro de los individuos, "algo" que los impulsa a comportarse de tal y cual manera. Esto significa que los investigadores deben medir siempre supuestos indicadores de motivación y no a ella en sí misma. En otras palabras, deben medir algún tipo de comportamiento, sean marcas en un papel, palabras habladas, o gestos significativos, y después hacer inferencias sobre características supuestas o variables latentes.

Se han usado otros términos para expresar más o menos las mismas ideas. Por ejemplo, Tolman (1951 pp. 115-129) llamó a los constructos variables intervinientes. Las *variables intervinientes* representan un término inventado para dar cuenta de procesos psicológicos no observables, internos, que a su vez dan cuenta de la conducta. Una variable interviniente es una variable "en la cabeza": No se le puede ver, oír o tocar. Se infiere a partir del comportamiento. La "hostilidad" se infiere de actos presumiblemente hostiles o agresivos. La "ansiedad" se infiere de la puntuación en una prueba, respuesta en la piel, frecuencia cardiaca y ciertas manipulaciones experimentales. Otro término es "constructo hipotético", cuyo significado es bastante similar al de variable latente, aunque con un poco menos de generalidad; no es necesario detenernos en él. Debemos mencionar, sin embar-

go, que "variable latente" parece ser una expresión más general y aplicable que "variable interviniente" y que "constructo hipotético", en tanto que puede usarse virtualmente para cualquier fenómeno que se presume influye o es influido por otro. En otras palabras "variable latente" puede utilizarse con fenómenos psicológicos, sociológicos y de otro tipo. "Variable latente" parece ser un vocablo más útil por su generalidad y también por que ahora es posible, en el enfoque de análisis de estructuras de covarianza, evaluar el efecto de las variables latentes entre sí y las llamadas variables manifiestas u observadas. Está discusión que parece muy abstracta después se concretará para ser, esperamos, significativa. Veremos entonces que la idea de las variables latentes y las relaciones entre ellas son extremo importantes, fructíferas y útiles, y ayudan a cambiar los enfoques fundamentales para afrontar los problemas de investigación.

Ejemplos de variables y definiciones operacionales

Hemos aportado una cantidad de constructos y definiciones operacionales. Para ilustrar y aclarar la discusión previa, en especial en lo referente a la diferencia entre variables experimentales y variables medidas, y entre constructos y variables definidas operacionalmente, presentamos diversos ejemplos. Si la definición es experimental se etiqueta con (E); si es una definición medida se señala como (M).

Las definiciones operacionales difieren en su grado de especificidad. Algunas están ligadas de manera estrecha a las observaciones. Definiciones de "pruebas", como "la inteligencia se define como una puntuación *x* en una prueba de inteligencia" son muy específicas. Una definición como "la frustración consiste en no alcanzar una meta" son más generales y requieren una mayor especificación para que sean medibles.

Clase social. "...dos o más grupos de personas que se cree que están en una posición social superior e inferior y que son así categorizados por los miembros de una comunidad" (M) (Warner y Lunt, 1941, p. 82). Para ser operacional, está definición ha de especificarse a partir de preguntas dirigidas a las creencias de la gente sobre las posiciones de otras personas. Se trata de una definición subjetiva de clase social. La clase social, o el estatus social, también es define de forma más objetiva a través de índices como ocupación, ingreso y educación, o por combinaciones de tales índices. Por ejemplo, "...transformamos la información acerca de educación, ocupación e ingresos de los padres de jóvenes de una liga juvenil en un índice de nivel socioeconómico (NSE), en el que las puntuaciones altas indican una avanzada educación, una ocupación prestigiosa y un ingreso desahogado. Calificaciones menores reflejan pobreza, educación incompleta y los trabajos más modestos" (M) (Herrnstein y Murray, 1996, p. 131).

Aprovechamiento (escolar, aritmético y en ortografía). El aprovechamiento se acostumbra definir operacionalmente a partir de una prueba estandarizada de aprovechamiento (por ejemplo la prueba de Iowa de Destrezas Básicas, la prueba de aprovechamiento o la prueba elemental de la batería de Evaluación de Infantil de Kaufman [K-ABC]), por medio del promedio o por el juicio del maestro. "El aprovechamiento del estudiante se midió por una puntuación combinada de pruebas de lectura y matemáticas" (M) (Peng y Wright, 1994). En ocasiones, el aprovechamiento se presenta en forma de una prueba de ejecución. Silverman (1993) examinó en un grupo de estudiantes dos habilidades del juego de voleibol: la prueba de servicio y la prueba del pase con antebrazo. En la primera, los estudiantes recibían una puntuación entre 0 y 4, en función de donde fuera colocado el balón servido. La prueba del pase con antebrazo consistía en hacer rebotar la pelota en el antebrazo. El criterio usado fue contar el número de veces en que un estudiante pudo pasar el balón más allá de una línea de dos metros y medio contra la pared en un periodo de un minuto (M). En algunos estudios educativos también se usa una definición operacional del

concepto *percepción del aprovechamiento del estudiante*. Aquí se pide a los estudiantes que se evalúen a sí mismos. La pregunta usada por Shoffner (1990) fue: "¿Qué clase de estudiante crees que eres?". Las opciones eran "estudiante de 9 y 10" "estudiante de 8" y "estudiante de 7 y 6" (M).

Aprovechamiento (desempeño académico). "Como resultado se obtuvieron las calificaciones de todos los estudiantes en todas las secciones y se usaron para determinar la categoría de cada estudiante participante en el estudio. Se calculó el rango porcentilar por sección para cada uno y se usó como la medida dependiente del aprovechamiento en el análisis final de los datos" (M) (Strom, Hocevar y Zimmer, 1990).

Motivación intrínseca se define operacionalmente por Horn, Berger *et al.* (1994) como "la cantidad acumulada de tiempo que cada estudiante juega con un patrón de bloques sin un sistema de reforzamiento" (M).

Popularidad. La popularidad con frecuencia se define operacionalmente por el número de elecciones sociométricas que un individuo recibe de otros (en su clase, grupo de juego, etcétera). Se le pregunta a los sujetos: "¿con quién te gustaría trabajar?", "¿con quién te gustaría jugar?" y otras cuestiones similares. Cada sujeto debe elegir a uno, dos o más individuos de su grupo con base en esas preguntas de criterios (M).

Compromiso con el trabajo "...el comportamiento de cada niño durante una lección fue evaluada cada seis segundos como apropiadamente comprometida o no. La puntuación del compromiso con el trabajo, por lección, fue el porcentaje de las unidades de seis segundos en que los niños fueron calificados como apropiadamente comprometidos" (M) (Kounin y Doyle, 1975).

Reforzamiento. Las definiciones de reforzamiento tienen diversas formas. La mayoría incluye, de una forma u otra, el principio de la recompensa. Sin embargo, se puede utilizar tanto el reforzamiento positivo como negativo. A continuación se presentan definiciones experimentales específicas.

En los siguientes 10 minutos cada opinión que el sujeto (S) enunció fue registrada por el experimentador (E) y reforzada. Para dos grupos, el E estuvo de acuerdo con la opinión enunciada al decir "sí tienes razón", "así es", o algo similar, o al sentir o sonreír en caso de que no pudiera interrumpir (E).

> ...se administraron al modelo y al niño de manera alternativa doce diferentes grupos de reactivos de historias... Ante cada uno de ellos, el modelo expresaba de forma consistente respuestas en forma de juicios opuestas a la orientación moral del niño... Y el experimentador reforzaba el comportamiento del modelo con respuestas de aprobación verbal tales como "muy bien", "está bien" y "qué bueno". El niño fue reforzado de manera parecida cada vez que adoptaba el tipo de juicios morales del modelo en respuesta a su propio grupo de reactivos [esto se denomina "reforzamiento social"] (E) (Bandura y MacDonald, 1994).

El maestro otorga al niño un reconocimiento verbal cada vez que exhibe el comportamiento deseado. Éstos son: atender a la instrucción, cumplir con el trabajo escolar y contestar en voz alta. El registro se hace cada 15 segundos (E) (Martens, Hiralall y Bradley, 1997).

Actitudes hacia el SIDA se define en una escala de 18 reactivos, cada uno con un formato de tipo Likert para reflejar diversas actitudes hacia los pacientes con SIDA. Algunos ejemplos de los reactivos son: "a la gente con SIDA no se le debería permitir usar los sanitarios públicos", y "debería ser obligatorio para todos los estadounidenses hacerse una prueba para SIDA" (M) (Lester, 1989).

Personalidad limítrofe. Comrey (1993) la define como la presencia de una baja puntuación en tres escalas de la Escala de Personalidad de Comrey: confianza *vs.* defensividad, conformidad social *vs.* rebeldía, y estabilidad emocional *vs.* neuroticismo.

Delincuencia laboral se define operacionalmente como una combinación de tres variables: número de accidentes imputables al sujeto, número de cartas de advertencia y número de suspensiones (M) (Hogan y Hogan, 1989).

Religiosidad se define como una puntuación en la Escala de Francis de Actitudes hacia la Cristiandad que consta de 24 reactivos con una escala de respuesta tipo Likert. Algunos reativos son: "Orar me ayuda mucho" y "Dios me guía para conducir una vida mejor" (M) (Gillings y Joseph, 1996). La religiosidad no debe confundirse con la preferencia religiosa. La religiosidad se refiere a la fuerza de la devoción a la religión que uno ha elegido.

Autoestima es una variable independiente manipulada en el estudio de Steele, Spencer y Lynch (1993). En este caso, se aplica a los sujetos una prueba de autoestima, pero cuando se les retroalimenta la información en el reporte de retroalimentación, que parece ser el oficial es ambigua. Se divide a los sujetos del mismo nivel de autoestima medida en tres grupos diferentes de retroalimentación: positiva, negativa y ausente. En la condición de retroalimentación positiva (autoestima positiva), se describe a los sujetos con enunciados tales como "pensamiento claro". Aquéllos en el grupo negativo (autoestima negativa) reciben adjetivos como "pasivos al actuar". Al grupo "sin retroalimentación" se les indica que su perfil de personalidad (autoestima) no estuvo listo por demoras en la calificación e interpretación (E). La mayoría de los estudios sobre autoestima usan una definición operacional medida. En el ejemplo anterior, Steele, Spencer y Lynch también usaron la Escala de Sentimientos de Inadecuación de Autoestima de Janis-Field (M). En otro ejemplo Luhtanen y Crocker (1992) definieron la autoestima colectiva como una puntuación en una escala de 16 reactivos tipo Likert en los que se solicitaba pensar a cerca de una variedad de grupos sociales y características de sus miembros, tales como género, religión, raza y grupo étnico (M).

La *raza* es por lo general una variable medida. Sin embargo en un estudio de Annis y Corenblum (1986), un experimentador (E) ya sea de raza blanca o india preguntó a 83 niños indios canadienses de nivel preescolar y primer año sobre preferencias raciales e identidad personal. El interés se centraba en averiguar si la raza del experimentador influía o no en las respuestas.

Soledad. Una definición de ésta es la puntuación en la Escala de Soledad de UCLA que incluye reactivos tales como: "nadie me conoce realmente bien" o "carezco de compañía". También se cuenta con la Escala de Deprivación y Soledad que presenta reactivos como: "experimento una sensación de vacío" o "no hay quien muestre un interés particular en mí" (M) (Oshagan y Allen, 1992).

Halo. Se han postulado muchas definiciones operacionales del efecto de halo. Balzer y Sulsky (1992) encontraron y resumieron 108 definiciones que ajustaron en 6 categorías. Una establece que halo es "...la variaza promedio intratasa o la desviación estándar de las evaluaciones". Otra puede ser: "Comparar las evaluaciones obtenidas con las proporcionadas por jueces expertos" (M).

Memoria: recuerdo y reconocimiento "...recuerdo significa pedir al participante que repita lo que recuerda de los reactivos que le fueron mostrados, y asignar un punto por cada uno que corresponda a la lista de estímulos inicial"(M) (Norman, 1976, p. 97). "La prueba de reconocimiento consistió en 62 frases presentadas a todos lo sujetos... que fueron instruidos para evaluar cada frase de acuerdo a su grado de confianza de que la oración hubiera sido presentada en el conjunto inicial" (M) (Richter y Seay, 1987).

Habilidades sociales. Pueden ser definidas operacionalmente como una puntuación en la Escala de Evaluación de Destrezas Sociales (Gresham y Elliot, 1990). Existe la posibilidad de contar con información del estudiante de su padre y del maestro. Se evalúan los comportamientos sociales en términos de frecuencia de ocurrencia y también de acuerdo a su nivel de importancia. Algunos reactivos incluyen: "se lleva bien con gente que es

diferente (maestro)", "se ofrece a ayudar a miembros de la familia con sus tareas (papá)", y, "cuestiono cortésmente las reglas que pueden ser injustas (estudiante)" (M).

Congraciamiento. Una de las muchas técnicas de manejo de impresión (véase Orpen, 1996; Gordon, 1996). Se define operacionalmente como una puntuación en la Escala de Kumar y Beyerlein (1991). Que comprende 25 reactivos tipo Likert diseñados para medir la frecuencia en que el subordinado, en una relación superior-subordinado usa tácticas para congraciarse (M). Strutton, Pelton y Lumpkin (1995) modificaron la Escala de Kumar-Beyerlein, para medir el congraciamiento entre el vendedor y cliente (M).

Feminismo. Se define por la puntuación en el Cuestionario de Actitudes hacia las mujeres. Este instrumento consta de 18 enunciados en los que el entrevistado registra su acuerdo en una escala de cinco puntos. Los reactivos incluyen: "los hombres han mantenido el poder por demasiado tiempo"; "los concursos de belleza son degradantes para la mujer"; "los niños de madres trabajadoras tienden a sufrir" (Wilson y Reading, 1989).

Valores. "Ordene las 10 metas de acuerdo a la importancia que tienen para usted: 1) Tener éxito financiero; 2) Ser querido; 3) Tener éxito en el ámbito familiar; 4) Ser capaz en lo intelectual; 5) Vivir de acuerdo a principios religiosos; 6) Ayudar a otros; 7) Ser normal, bien ajustado; 8) Cooperar con los demás; 9) Trabajar con detalle; 10) Alcanzar el éxito laboral" (M) (Newcomb, 1978).

Democracia (democracia política). "El índice (de democracia política) consiste en tres indicadores de soberanía popular y tres de libertades políticas. Las medidas de soberanía popular son: 1) Elecciones limpias, 2) Selección ejecutiva efectiva, y 3) Selección legislativa. Los indicadores de libertades políticas son: 4) Libertad de prensa, 5) Libertad para que la oposición se asocie y 6) Sanciones gubernamentales" (M). Bollen (1979) proporciona detalles operacionales de seis indicadores sociales en un apéndice (pp. 585-586). Éste es un ejemplo particularmente bueno de la definición operacional de un concepto complejo. Más aún, constituye una excelente descripción de los ingredientes de la democracia.

Los beneficios del pensamiento operacional han sido grandiosos. De hecho el operacionalismo ha sido y es uno de los movimientos más significativos e importantes de nuestro tiempo. El operacionalismo extremo, por supuesto, puede ser peligroso porque oscurece el reconocimiento de la importancia de los constructos y definiciones constitutivas en la ciencia del comportamiento y porque puede restringir la investigación a problemas triviales. Sin embargo, hay poca duda de que sea una sana influencia. Resulta una clave indispensable para alcanzar la objetividad (sin la cual no hay ciencia), en tanto que demanda que las observaciones sean públicas y replicables, lo que ayuda a colocar las actividades de investigación más allá de los investigadores y sus predilecciones. Y, como dijo Underwood (1957, p. 53) en su texto clásico sobre investigación psicológica:

> Yo diría que el pensamiento operacional hace mejores científicos. El operacionalista se ve forzado a sacudir y aclarar sus conceptos empíricos... el operacionalismo facilita la comunicación entre científicos ya que el significado de los conceptos así definidos no es sujeto fácilmente a una mala interpretación.

RESUMEN DEL CAPÍTULO

1. Un *concepto* es una expresión de una abstracción formada a partir de la generalización de un particular, por ejemplo, peso. Esta expresión se deriva de observaciones de ciertos comportamientos o acciones.
2. Un *constructo* es un concepto que se ha formulado para ser usado en la ciencia. Se usa en esquemas teóricos y se define de tal manera que sea susceptible de ser observado y medido.

3. Una *variable* se define como una propiedad que puede tomar diferentes valores; es un símbolo al que se le asignan valores.
4. Los constructos y las palabras pueden ser definidas por
   a) otras palabras o conceptos.
   b) descripción de una acción o conducta implícita o explícita.
5. Una *definición constitutiva* se da cuando los constructos están definidos por otros constructos.
6. Una *definición operacional* se presenta cuando se aporta el significado al especificar las actividades u operaciones necesarias para medir y evaluar el constructo. Las definiciones operacionales sólo pueden dar un significado limitado al constructo; no pueden describir completo a un constructo o variable. Hay dos tipos de definiciones operacionales:
   a) de medida —nos dice cómo será medida la variable o constructo—.
   b) experimental —explica los detalles de cómo el experimentador manipula la variable (constructo)—.
7. Tipos de variables
   a) La *independiente* varía y es la causa supuesta de otra variable, la dependiente. En un experimento, constituye la variable manipulada, es la que está bajo el control del experimentador. En un estudio no experimental, es la variable que tiene un efecto lógico en la variable dependiente.
   b) El efecto de la variable *dependiente* se altera de forma concomitante con los cambios o variaciones en la variable independiente.
   c) Una variable *activa* se manipula. Manipulación significa que el experimentador tiene control sobre cómo cambian los valores.
   d) Una variable *atributiva* se mide y no puede ser manipulada, es decir, es aquella donde el experimentador no tiene control sobre los valores de la variable.
   e) Una variable *continua* es capaz de asumir un grupo ordenado de valores dentro de cierto rango. Entre dos valores hay un número infinito de otros valores. Esta variable refleja por lo menos una categoría ordinal.
   f) Las variables *categóricas* pertenecen a una clase de medición donde los objetos se asignan a subclases o a subgrupos, diferenciados y que no se traslapan. Se considera que todos los elementos de una misma categoría tienen la misma característica o características.
   g) Las variables *latentes* son entidades no observables y se asume que subyacen a las variables observadas.
   h) Las variables *intervinientes* son constructos que dan cuenta de procesos psicológicos internos no observables que explican el comportamiento. No pueden ser vistas pero se infieren a partir del comportamiento.

SUGERENCIAS DE ESTUDIO

1. Escriba las definiciones operacionales para 5 o 6 de los siguientes constructos. Cuando le sea posible, escriba dos definiciones: una experimental y una definición de medida.

reforzamiento	poder de castigo
aprovechamiento	habilidad lectora
bajo rendimiento	necesidades
liderazgo	interés

transferencia del entrenamiento	delincuencia
nivel de aspiración	necesidad de afiliación
conflicto organizacional	conformidad
preferencia política	satisfacción marital

Alguno de estos conceptos o variables —por ejemplo, necesidades y transferencia del entrenamiento— pueden resultar difíciles de definir operacionalmente. ¿Por qué?

2. ¿Podría alguna de las variables del inciso anterior ser tanto variable independiente como variable dependiente? ¿Cuáles?
3. Resulta instructivo y útil para los especialistas leer acerca de otros campos distintos del propio. Esto es en particular cierto para los estudiantes de investigación del comportamiento. Se sugiere que el estudiante de un campo en particular lea dos o tres estudios de investigación en una de las mejores revistas de otra disciplina. Si usted está en psicología lea una revista de sociología, por ejemplo, la *American Sociological Review*. Si usted está inmerso en el campo de la educación o la sociología, lea una revista de psicología como el *Journal of Personality y Social Psychology* o el *Journal of Experimental Psychology*. Los alumnos que no pertenecen al área de educación, pueden hojear el *Journal of Educational Psychology* o el *American Educational Research Journal*. Al leer, tome nota de las variables y compárelas con las de su propio campo. ¿Son primariamente variables activas o variables atributo? Observe, por ejemplo, qué variables psicológicas son más "activas" que las sociológicas. ¿Qué implican las variables de una disciplina para su investigación?
4. La lectura de los siguientes artículos será útil para entender y desarrollar definiciones operacionales.

Kinnier, R.T. (1995). A reconceptualization of values clarification: Values conflict resolution. *Journal of Counseling and Development, 74(1)*, 18-24.

Lego, S. (1988). Multiple disorder: An interpersonal approach to etiology, treatment and nursing care. *Archives of Psychiatric Nursing, 2(4)*, 231-235.

Lobel, M. (1994). Conceptualizations, measurement, and effects of prenatal maternal stress on birth outcomes. *Journal of Behavioral Medicine, 17(3)*, 225-272.

Navathe, P.D. & Singh, B. (1994). An operational definition for spatial disorientation. *Aviation, Space & Environmental Medicine, 65(12)*, 1153-1155.

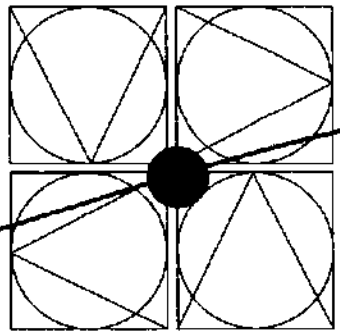
Sun, K. (1955). The definition of race. *American Psychologist, 50(1)*, 43-44.

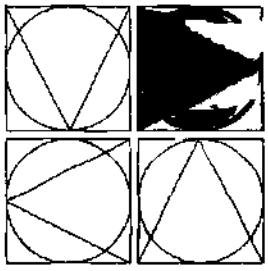
Talaga, J.A. & Beehr, T.A. (1995). Are the gender differences in predicting retirement decisions? *Journal of Applied Psychology, 80(1)*, 16-28.

Woods, D.W. Miltenberger, R.G. & Flach, A.D. (1996). Habits, tics and stuttering: Prevalence and relation to anxiety and somatic awareness. *Behavior Modification, 20(2)*, 216-225.

Parte Dos

Conjuntos, relaciones y varianza





CAPÍTULO 4

CONJUNTOS

- SUBCONJUNTOS
- OPERACIONES DE CONJUNTOS
- CONJUNTOS UNIVERSALES Y VACÍOS; LA NEGACIÓN DEL CONJUNTO
- DIAGRAMAS DE CONJUNTOS
- OPERACIONES CON MÁS DE DOS CONJUNTOS
- PARTICIONES Y PARTICIONES CRUZADAS
- NIVELES DEL DISCURSO

El concepto de "conjunto" es una de las ideas matemáticas más poderosas y útiles para comprender los aspectos metodológicos de la investigación. Los conjuntos y sus elementos son la materia prima que subyace a la forma en que opera la matemática. Aun si no estamos conscientes de ello, los conjuntos y la teoría de conjuntos son fundamentos de nuestro pensamiento descriptivo lógico y analítico así como de su operación. Virtualmente, constituyen la base de todo lo que se presenta en este libro y son los cimientos sobre los cuales erigimos complejos análisis numéricos, categóricos y estadísticos, incluso cuando no constituyen los fundamentos explícitos de nuestro pensamiento y trabajo. Por ejemplo, la teoría de conjuntos proporciona una definición no ambigua de relaciones, nos ayuda a aproximarnos y a entender la probabilidad y el muestreo, y es prima hermana de la lógica. También nos ayuda a entender el muy importante tema de las categorías y la categorización de los objetos del mundo. Más aún, el pensamiento de conjuntos puede apoyarnos para comprender ese difícil problema de la comunicación humana: la confusión causada por mezclar diferentes niveles del discurso.

La ciencia trabaja básicamente con conceptos de grupo, clase o conjunto. Cuando los científicos discuten sobre eventos u objetos individuales, los consideran como miembros de conjuntos de objetos. Pero esto es cierto también en el discurso humano en general. Decimos "ganso", pero la palabra *ganso* no tiene sentido sin el concepto de un grupo de tales individuos llamado "gansos". Cuando hablamos acerca de un niño y de sus problemas, resulta inevitable hablar de los grupos, clases o conjuntos de objetos a los que el niño pertenece. Esto incluiría a un niño de 7 años (primer conjunto), de segundo grado (segun-

do conjunto), brillante (tercer conjunto), saludable (cuarto conjunto) y varón (quinto conjunto).

De acuerdo con Farlow (1988) y Smith (1992), un *conjunto* es una colección bien definida de objetos. Un conjunto se encuentra bien definido cuando no hay duda sobre si un objeto dado pertenece o no al conjunto. Palabras como clase, escuela, familia, manada y grupo, indican conjuntos. Hay dos formas para definirlos: 1) a través de un listado de todos los miembros del conjunto, y 2) con una regla para determinar cuáles objetos pertenecen o no al conjunto. Llamaremos al 1) una definición de "lista" y a 2) una definición de "regla". En investigación se usa por lo general la definición de regla, aunque hay casos donde se enlista a todos los miembros de un conjunto por escrito o mentalmente. Por ejemplo, supongamos que estudiamos la relación entre la conducta del votante y la preferencia política. En *Estados Unidos la preferencia política* puede definirse como estar registrado como republicano o demócrata. Así, tenemos entonces un conjunto amplio de todas las personas con preferencias políticas en dos subconjuntos más pequeños: el subconjunto de los republicanos y el de los demócratas. Ésta es una definición de conjuntos a partir de una regla. Por supuesto, podríamos tener una lista de todos los demócratas y republicanos registrados para definir ambos subconjuntos, pero con frecuencia es difícil o imposible, además, resulta innecesario. La regla, en general, basta. Podría frasearse así: un republicano es cualquier persona que esté registrada en el partido republicano. Otra regla podría ser: un republicano es cualquier persona que diga que es republicano.

Subconjuntos

Un *subconjunto* de un conjunto es un conjunto que resulta de seleccionar conjuntos de un conjunto original. Cada subconjunto de un conjunto es parte del conjunto original. De forma más sucinta y precisa, el conjunto B es un subconjunto de un conjunto A siempre que todos los elementos de B son elementos de A (Kershner & Wilcox 1974). Designamos a los conjuntos con letras mayúsculas: A, B, K, L, X, Y, y así sucesivamente. Si B es un subconjunto de A, nosotros escribimos $B \subset A$, que significa "B es subconjunto de A", "B está contenido en A" o "todos los miembros de B son también miembros de A".

Cuando se muestrea una población, las muestras resultantes son subconjuntos de la población. Suponga que un investigador muestrea cuatro clases de 6o. grado, de todos los grupos de 6o. en una escuela grande. Las cuatro clases forman un subconjunto de la población de todas las clases de 6o. grado. Cada una de las cuatro clases de la muestra puede considerarse también subconjunto de las cuatro clases —y también de la población total de los grupos—. Todos los estudiantes de las cuatro clases pueden dividirse en dos subconjuntos: niños y niñas. Cuando un investigador fracciona o parte una población o una muestra en dos o más grupos, los subconjuntos se crean por medio de una "regla" o criterio para hacerlo. Los ejemplos son numerosos: dividir preferencia religiosa en protestante, católica, judía; inteligencia en alta y baja; etcétera. También podemos observarlo en condiciones experimentales: el modelo clásico de grupo experimental y grupo control es una idea conjunto-subconjunto. Hay individuos que se asignan al grupo experimental, que constituye un subconjunto de toda la muestra. Todos los demás individuos del experimento (los del grupo control) también forman un subconjunto.

Operaciones de conjuntos

Hay dos operaciones básicas de conjuntos: *intersección* y *unión*. Una operación consiste tan sólo en "hacer algo para". En aritmética sumamos, restamos, multiplicamos y dividimos.

En el ámbito de los conjuntos, "intersecamos" y "unimos". También los "negamos". Cuando tratamos con conjuntos, hay operadores lógicos involucrados. Para la intersección, el operador lógico es "y"; para la unión, el operador lógico apropiado es "o" y para la negación, el operador es "no". Si se desea saber más de operadores lógicos sugerimos leer a Udolf (1973).

La *intersección* es el traslape de dos o más conjuntos; consiste en los elementos compartidos en común por dos o más conjuntos. El símbolo para la intersección es $\cap$ (se lee "intersección") la intersección de los conjuntos A y B se escribe $A \cap B$, y $A \cap B$ es en sí misma un conjunto. Con mayor precisión, es el conjunto que contiene aquellos elementos de A y B que pertenecen a ambos A y B. La intersección también se escribe $A * B$, o simplemente AB.

Sea $A = \{0, 1, 2, 3\}$; sea $B = \{2, 3, 4, 5\}$. (Note que las llaves "{}" se usan para simbolizar conjuntos.) Entonces $A \cap B = \{2, 3\}$, como se muestra en la figura 4.1. $A \cap B$, o $\{2, 3\}$, es un nuevo conjunto compuesto por los elementos *comunes* a ambos conjuntos. Observe que $A \cap B$ también indica la *relación* entre los conjuntos —los elementos compartidos por A y B—.

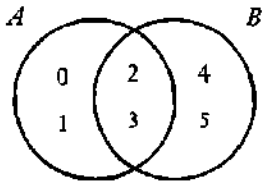
La *unión* de dos conjuntos se escribe $A \cup B$ que es un conjunto que contiene a todos los miembros de A y a todos los de B. Los matemáticos definen $A \cup B$ como un conjunto que contiene aquellos elementos que pertenecen a A o a B, o a ambos. En otras palabras, "sumamos" los elementos de A a aquellos de B para formar el nuevo conjunto $A \cup B$. Tomemos el ejemplo de la figura 4.1. A incluye a 0, 1, 2 y 3; B incluye a 2, 3, 4 y 5. $A \cup B = \{0, 1, 2, 3, 4, 5\}$. La unión de A y B en la figura 4.1 se indica por toda el área de ambos círculos. Observe que no contamos a los miembros de $A \cap B$ {2, 3} dos veces.

Algunos ejemplos de unión en investigación podrían ser agrupar juntos a hombres y mujeres, $M \cup F$ o a los republicanos y demócratas, $R \cup D$. Sea A todos los niños de escuelas primarias y B, todos los niños de escuelas secundarias de X distrito escolar. Entonces, $A \cup B$ es el conjunto de todos los niños de las escuelas en el distrito.

Conjuntos universales y vacíos; la negación del conjunto

El conjunto universal, que se escribe U, es el conjunto de todos los elementos bajo discusión. Puede llamarse *universo del discurso* o *nivel del discurso*. (Equivale al término *población* y *universo* en teoría de muestreo.) Esto implica que limitamos nuestra discusión a un conjunto definido de elementos —todos ellos— pertenecientes a la clase determinada, U. Si estudiamos los determinantes del aprovechamiento en la escuela primaria, por ejemplo, podemos definir U como todos los alumnos en grados 1o. al 6o. Podemos definir U, de forma alternativa, como las puntuaciones en una prueba de aprovechamiento de estos mismos alumnos. Los subconjuntos de U que pudieran estudiarse por separado, podrían ser las puntuaciones de los alumnos de 1er. grado, las de los estudiantes de 2o. grado, etcétera.

FIGURA 4.1



U puede ser grande o pequeño. Si regresamos al ejemplo de la figura 4.1, $A = \{0, 1, 2, 3\}$ y $B = \{2, 3, 4, 5\}$ y si $A \cup B = U$, entonces $U = \{0, 1, 2, 3, 4, 5\}$. Aquí U es muy pequeño. Supongamos que $A = \{$Juana, María, Felicia, Beatriz$\}$ y $B = \{$Tomás, Juan, Pablo$\}$. Si sólo hablamos de estos individuos, entonces $U = \{$Juana, María, Felicia, Beatriz, Tomás, Juan, Pablo$\}$ y por supuesto, $U = A \cup B$. Éste es otro ejemplo de un pequeño U. En investigación, los U son con frecuencia grandes. Si muestreamos las escuelas de un condado grande, entonces U sería todas las escuelas del condado, es decir, un U relativamente grande. U podría también estar constituido por todos los niños o todos los maestros en estas escuelas, un U aún mayor.

En investigación, es importante conocer el U que uno estudia. La ambigüedad en la definición de U puede llevarnos a conclusiones erróneas. Se sabe, por ejemplo, que las clases sociales difieren en cuanto a la incidencia de neurosis y psicosis (Murphy, Olivier, Monson y Sobol, 1991). Si estudiáramos determinantes supuestos de enfermedad mental y trabajamos sólo con sujetos de clase media, nuestras conclusiones, por supuesto, se limitarían tan sólo a ese sector social. Es fácil generalizar a todas las personas, pero esa práctica puede constituir un gran error. En ese caso, habríamos generalizado a todas las personas, U, cuando de hecho apenas hemos estudiado las relaciones en U_1, la clase media. Es muy posible, incluso probable, que las relaciones sean diferentes en U_2, la clase trabajadora.

El *conjunto vacío* es un conjunto sin miembros y se denota E. También se le llama conjunto *nulo*. Aunque pudiera parecer peculiar al estudiante que nos preocupemos por los conjuntos sin miembros, el concepto es muy útil y aun indispensable. Con ellos podemos transmitir ciertas ideas de una forma económica y sin ambigüedades. Para indicar que no hay relación entre dos conjuntos de datos, por ejemplo, podemos escribir la ecuación del conjunto $A \cap B = E$ que simplemente indica que la intersección de los conjuntos A y B está vacía, es decir, que no hay miembros de A que pertenezcan a B, y viceversa.

Sea $A = \{1, 2, 3\}$; sea $B = \{4, 5, 6\}$. Entonces, $A \cap B = E$. Es evidente que no hay miembros comunes a A y B. El conjunto de posibilidades de que tanto el candidato presidencial demócrata como el republicano ganen la elección nacional es un conjunto vacío (E). El conjunto de ocurrencias de lluvia sin nubes está vacío (E). El conjunto vacío es, pues, otra forma de expresar la falsedad de una proposición. En este caso podemos decir que el enunciado "lluvia sin nubes" es falso. En lenguaje de conjuntos esto puede expresarse $P \cap Q = E$ donde P es igual al conjunto de todas las ocurrencias de lluvia, Q es igual al conjunto de ocurrencias de nubes, y $\sim Q$ equivale al conjunto de todas las ocurrencias de no nubes.

La *negación o complemento* del conjunto A se escribe $\sim A$. Significa todos los miembros de U que no están en A. Si asumimos que A es igual a todos los hombres, cuando U equivale a todos los seres humanos, entonces $\sim A$ es igual a todas las mujeres (no-hombres). La dicotomización simple parece ser una base fundamental del pensamiento humano. La categorización es necesaria para pensar: uno debe, en el nivel más elemental, separar los objetos en aquellos que pertenecen a cierto conjunto y aquellos que no. Debemos distinguir en humanos y no-humanos, entre yo y no-yo, temprano y no-temprano, bueno y no-bueno.

FIGURA 4.2

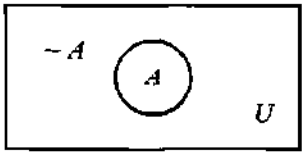
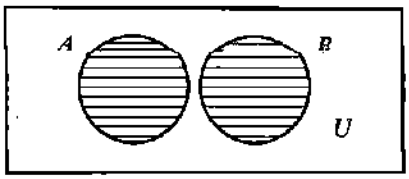


FIGURA 4.3



$A \cup B$

Si $U = \{0, 1, 2, 3, 4\}$ y $A = \{0, 1\}$ entonces $\sim A = \{2, 3, 4\}$. A y $\sim A$ son, por supuesto, subconjuntos de U. Una propiedad importante de los conjuntos y su negación se expresa en la ecuación de conjuntos: $A \cup \sim A = U$. Observe también que $A \cap \sim A = E$.

Diagramas de conjuntos

Ahora podemos recapitular e ilustrar las ideas básicas de los conjuntos que hemos presentado, a través de diagramas. Es posible representarlos con diversas figuras, pero los rectángulos y círculos son los más comunes. Se han adaptado de un sistema inventado por John Venn, un lógico del siglo XIX. En este libro usaremos rectángulos, círculos y óvalos. Observe la figura 4.2 donde U se representa con un rectángulo. Todos los elementos del universo bajo discusión están en U. Todos los miembros de U que no están en A forman otro subconjunto de U: $\sim A$. Observe una vez más que $A \cup \sim A = U$. Note también que $A \cap \sim A = E$; esto es, no hay elementos comunes a ambos, A y $\sim A$.

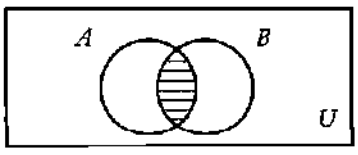
A continuación representamos (figura 4.3) dos conjuntos, A y B, ambos subconjuntos de U. A partir del diagrama puede observarse que $A \cap B = E$. Adoptamos una convención: cuando deseamos indicar un conjunto o un subconjunto, lo sombreamos ya sea de manera horizontal, vertical o diagonal. El conjunto $A \cup B$ ha sido sombreado en la figura 4.3.

La intersección, quizá el concepto más importante desde el punto de vista de este libro, se indica por la porción sombreada en la figura 4.4 y puede expresarse por la ecuación $A \cap B \neq E$; la intersección de los conjuntos A y B *no* está vacía.

Cuando dos conjuntos, A y B, son iguales, tienen los mismos elementos o miembros. El diagrama de Venn mostraría dos círculos congruentes en U. En efecto, sólo se vería un círculo. Cuando $A = B$, entonces $A \cap B = A \cup B = A = B$.

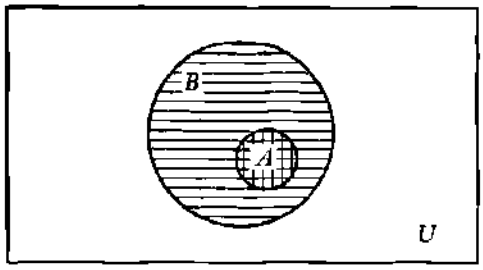
Trazamos el diagrama $A \subset B$; A es un subconjunto de B, en la figura 4.5. B se ha sombreado de forma horizontal, y A de manera vertical. Observe que $A \cup B = B$ (área sombreada completa) y $A \cap B = A$ (el área sombreada tanto horizontal como verticalmente). Todos los miembros de A están también en B, o todas las a son también b, si asumimos que a es igual a cualquier miembro de A y b es igual a cualquier miembro de B.

FIGURA 4.4



$A \cap B$

FIGURA 4.5



$A \subset B$

Operaciones con más de dos conjuntos

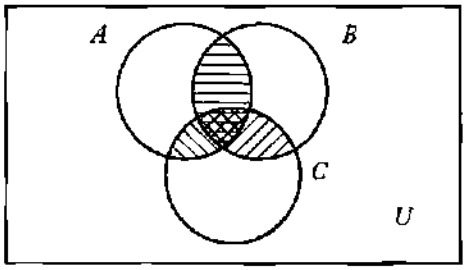
Las operaciones de conjunto no se limitan a dos subconjuntos de U. Sean A, B y C tres subconjuntos de U. Suponga que la intersección de estos tres subconjuntos de U no está vacía, como se muestra en la figura 4.6. El área con triple sombreado muestra $A \cap B \cap C$. Hay cuatro intersecciones, cada una sombreada de forma diferente: $A \cap B$, $A \cap C$, $B \cap C$ y $A \cap B \cap C$.

Aunque se pueden diagramar cuatro o más conjuntos, tales diagramas resultan difíciles de manejar, dibujar y revisar. No hay razón, sin embargo, para que las operaciones de intersección y unión no se apliquen simbólicamente a cuatro o más conjuntos.

Particiones y particiones cruzadas

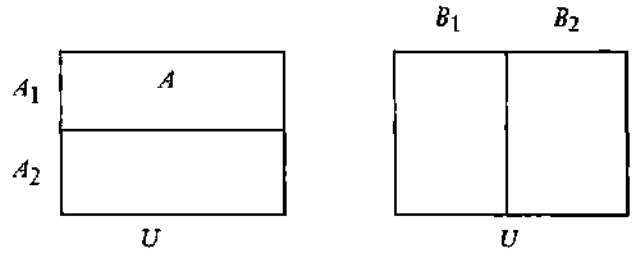
Nuestra discusión de conjuntos ha sido abstracta y quizás un poco monótona. Dejemos la discusión para examinar un aspecto de la teoría de conjuntos de gran importancia en la clarificación de los principios de categorización, análisis y diseño de investigación: la partición. U puede fraccionarse (partirse) en subconjuntos que no se intersequen y que agoten a U por completo. Cuando esto se hace, el proceso se denomina *partición*. Dicho formalmente, la partición fracciona un conjunto universal en subconjuntos *inconexos* y *exhaustivos* del conjunto universal.

FIGURA 4.6



$A \cap B \cap C$

FIGURA 4.7



Sea U un universo, y A y B los subconjuntos de U que son particiones. Llamamos a los subconjuntos de A: $A_1, A_2 \ldots, A_k$ y de B: $B_1 B_2 \ldots, B_m$. Las particiones por lo general se separan por corchetes mientras que los conjuntos y subconjuntos están marcados por paréntesis o llaves. Ahora, $[A_1 A_2]$ y $[B_1 B_2]$, por ejemplo son particiones si:

$$A_1 \cup A_2 = U \text{ y } A_1 \cap A_2 = E$$
$$B_1 \cup B_2 = U \text{ y } B_1 \cap B_2 = E$$

Los diagramas lo aclaran: la partición de U (representado por un rectángulo) por separado en los subconjuntos A_1 y A_2 y en B_1 y B_2 se muestra en la figura 4.7. Ambas particiones se han realizado en el mismo U. Hemos presentado algunos ejemplos de tales particiones: hombre-mujer, clase media-clase trabajadora, ingreso alto-ingreso bajo, demócrata-republicano, aprobado-reprobado, etcétera. Algunas son verdaderas dicotomías; otras no.

Es posible unir ambas particiones en una partición cruzada. Una *partición cruzada* es una nueva fracción que procede de sucesivas particiones del mismo conjunto U al formar todos los subconjuntos de la forma $A \cap B$. En otras palabras realizamos la partición A, y después la partición B en el mismo U, o en el mismo cuadrado, como se muestra en la figura 4.8. Cada casilla de la partición es una intersección de los subconjuntos A y B. Encontraremos en un capítulo posterior que tal partición cruzada es muy importante en el diseño de investigación y en el análisis de datos.

En forma anticipada a lo que se verá más adelante, he aquí un ejemplo de investigación de una partición cruzada. Dichos ejemplos se denominan tablas cruzadas o de contingencia. *Las tablas cruzadas* proveen la forma más elemental de mostrar una relación entre dos variables. El ejemplo es de un estudio de Miller y Swanson (1960), sobre prácticas de crianza infantil. Una de las tablas que usaron es una tabla cruzada cuyas variables son clase social (clase media y clase trabajadora) y destete (temprano y tardío). Los datos convertidos en porcentajes se muestran en la tabla 4.1. Las frecuencias reportadas por Miller y

FIGURA 4.8

		B	
		B_1	B_2
A	A_1	$A_1 \cap B_1$	$A_1 \cap B_2$
	A_2	$A_2 \cap B_1$	$A_2 \cap B_2$

TABLA 4.1 *Tabulación cruzada: relación entre clase social y destete (estudio de Miller y Swanson)*

		Destete	
		Temprano (B_1)	*Tardío* (B_2)
Clase social	*Clase media* (A_1)	60% (33)	40% (22)
	Clase trabajadora (A_2)	35% (17)	65% (31)

Swanson se pueden observar en el ángulo inferior derecho de cada casilla. Es evidente que existe una relación entre clase social y destete: las madres de clase social media muestran una tendencia a destetar a sus hijos más temprano que las de clase trabajadora. Se satisfacen las dos condiciones de inconexión y exhaustividad. La intersección de dos celdas cualesquiera está vacía. Por ejemplo, $(A_1 \cap B_1) \cap (A_1 \cap B_2) \cap (A_2 \cap B_1) \cap (A_2 \cap B_2) = E$. Las casillas agotan todos los casos: $(A_1 \cap B_1) \cup (A_1 \cap B_2) \cup (A_2 \cap B_1) \cup (A_2 \cap B_2) = U$.

La partición, por supuesto, va más allá de dos fracciones. En lugar de dicotomías, podemos tener politomías; en lugar de éxito-fracaso, por ejemplo, podemos tener éxito-éxito parcial-fracaso. Teóricamente una variable puede ser fraccionada en cualesquier número de subconjuntos, aunque por lo general hay limitaciones prácticas. Tampoco existe limitación teórica en el número de variables en una partición cruzada, pero las consideraciones prácticas por lo común limitan el número a tres o cuatro. El estudio de Foster, Dingman, Muscolino y Jankowski (1996) demuestra cómo se parte una variable en tres categorías. Su investigación es acerca de las decisiones relativas a contratación de empleados. Cada participante actuaba como gerente de recursos humanos y debía revisar tres currícula. El nombre en el documento determinaba el sexo del candidato. Los participantes debían recomendar un solicitante para la posición vacante. Los investigadores desarrollaron dos paquetes diferentes. En uno, un candidato masculino es el más calificado y la mujer es la de menores atributos. El segundo paquete es el reverso del anterior: la persona más calificada era mujer y el hombre era el menos apto. Había un tercer currículum de una persona cuyo sexo no podía ser determinado por el nombre. El cuadro que Foster y sus colaboradores presentan es una tabla cruzada en la que las variables son género del revisor/decisor (hombre y mujer) y tipo de currícula (altamente calificado masculino, menos apto masculino, altamente calificado femenino, menos apto femenino, altamente calificado sexo desconocido y menos apto sexo desconocido). Estos datos, convertidos en porcentajes, se muestran en la tabla 4.2. Las frecuencias obtenidas por Foster y colaboradores se presentan en la esquina inferior derecha de cada casilla, e ilustran una relación entre el género del revisor y el género del candidato. Las mujeres revisoras (quienes tomaban la decisión) tendían a seleccionar candidatos mujeres al hacer su recomendación. Los hombres revisores tendían a seleccionar candidatos masculinos aun cuando el candidato

TABLA 4.2 *Tabla de participación cruzada: relación entre género del revisor y calificación del solicitante (estudio de Foster et al.)*

Género del revisor	Currículum A Altamente calificado	Currículum B Menos calificado	Currículum C Menos calificado
(*Paquete 1*)	Jimena	Reme	Jorge
Femenino	50% (12)	33% (8)	17% (4)
Masculino	41% (7)	24% (4)	35% (6)
(*Paquete 2*)	Andrés	Michel	Carmen
Femenino	45% (9)	5% (1)	50% (10)
Masculino	53% (10)	26% (5)	21% (4)

femenino fuera superior en calificación. En un capítulo posterior se describirá con más detalle esta partición de variables.

Niveles del discurso

Cuando hablamos de cualquier asunto lo hacemos en un contexto o marco de referencia. Las expresiones, contexto y marco de referencia se relacionan de manera estrecha con *U*, el universo del discurso. Este último debe ser capaz de incluir cualesquier objetos de los que hablamos. Si vamos a otro *U* (otro nivel de discurso), el nuevo nivel no incluirá todos los objetos. De hecho, puede no incluir a objeto alguno. Si hablamos acerca de gente, por ejemplo, nosotros no hablamos —o mejor dicho, no deberíamos hablar— acerca de aves y sus costumbres a menos que de alguna forma relacionemos a las aves y sus costumbres con la gente, y que sea muy claro que esto es lo que hacemos. Hay dos niveles de discurso o universos (*U*) de discurso aquí: gente y aves. Cuando discutimos las implicaciones democráticas de la segregación, no debemos abordar de forma abrupta las preferencias religiosas, a menos que de alguna forma las relacionemos. Si lo hacemos, perderemos nuestro universo original de discurso, o no podremos asignar los objetos de un nivel (quizá religión) al otro nivel (la educación de los niños afroamericanos).

Para matizar esta presentación, cambiemos nuestro nivel de discurso a la música y el juicio y el entendimiento de diferentes géneros musicales. Una de las grandes dificultades al escuchar música moderna es que el sistema clásico de reglas que nuestros oídos han aprendido no se adecua al tipo de música de compositores como Bartók, Schöenberg o Ives. Uno tiene menos dificultades con Bartók y mucho más con Schöenberg e Ives porque el primero mantiene más bases clásicas que los otros dos. Examinemos la "Sonata Concordia" de Ives, en verdad un buen trabajo. La primera vez uno se sorprende por la aparente cacofonía y carencia de estructura. Después de escucharla varias veces, uno empieza a eliminar los juicios con base en el marco de referencia de la música clásica y a escuchar la belleza, significado y estructura del trabajo. El universo del discurso musical de Ives es simplemente muy diferente del universo del discurso clásico y resulta en extremo difícil desplazarse del *U* clásico al *U* de Ives. Algunas personas no pueden o incluso rechazan intentar este cambio. Encuentran la música de Ives extraña, inclusive repugnan-

te. Son incapaces de sacudirse el nivel de discurso de los conceptos estéticos y de juicio de la música clásica para hacer este desplazamiento.[1]

En la investigación debemos ser cuidadosos de no mezclar o desplazar nuestros niveles de discurso, y hacerlo sólo de manera consciente y con conocimiento. Pensar en términos de conjuntos nos ayuda a evitar estas mezclas y cambios. Como ejemplo extremo, supongamos que un investigador decide estudiar el entrenamiento para el control de esfínteres, el autoritarismo, las aptitudes musicales, la creatividad, la inteligencia, el aprovechamiento en lectura y el aprovechamiento escolar general de jóvenes de 3er. año de educación media. Aunque es concebible que se pueda extraer algún tipo de relaciones de este arreglo de variables, es más factible que se trate de una confusión intelectual. A cualquier costo, recuerde los conjuntos. Pregúntese: "Los objetos de los que discuto ¿pertenecen a un conjunto o conjuntos de mi presente discusión?" Si es así entonces usted está en un nivel de discurso. En caso contrario entra en la discusión otro nivel de discurso, otro conjunto, o conjunto de conjuntos. Si esto ocurre sin que se dé cuenta, el resultado es una confusión. En pocas palabras preguntemos: "¿Cuáles son U y los subconjuntos de U?"

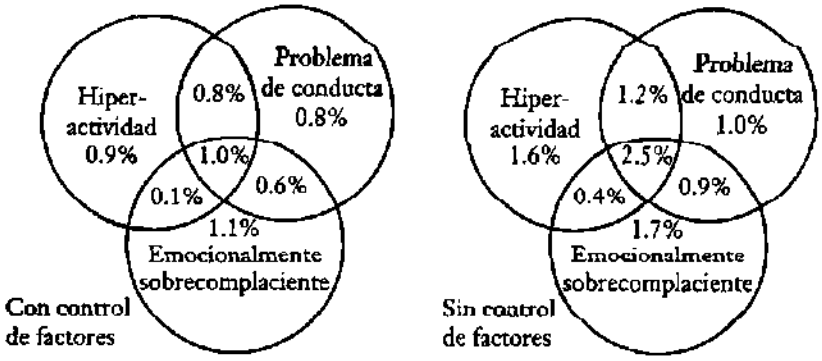
La investigación requiere definiciones precisas de los conjuntos universales. *Precisas* significa aportar una regla clara que nos diga cuándo un objeto es o no miembro de U. De manera similar, defina los subconjuntos de U y los subconjuntos de los subconjuntos de U. Si los objetos de U son personas, entonces no puede tener un subconjunto con objetos que no sean personas. (Aunque es posible tener un conjunto A de personas y el conjunto $\sim A$ de no-personas, esto nos lleva lógicamente a que U son personas. "No-personas" es, en este caso, un subconjunto de "personas", por definición o por convención.)

La idea de conjunto es fundamental en el pensamiento humano, en tanto que es probable que la mayor parte del pensamiento dependa de asignar cosas a categorías y de etiquetar esas categorías (véase Ross y Murphy, 1996; Smith, 1995). Lo que hacemos es agrupar clases de objetos —cosas, personas, eventos, fenómenos en general— y nombrar dichas clases. Tales nombres se convierten entonces en conceptos, etiquetas que no necesitamos aprender de nuevo y que pueden usarse para un pensamiento eficiente.

La teoría de conjuntos también constituye un instrumento general y ampliamente aplicable del pensamiento analítico y conceptual. Sus aplicaciones más importantes pertinentes a la metodología de la investigación son tal vez al estudio de las relaciones, la lógica, el muestreo, la probabilidad, la medición y el análisis de datos (véase Curtis, 1985; Hays, 1994). Sin embargo, la teoría de conjuntos puede aplicarse también a otras áreas y problemas no considerados técnicos en el sentido en que la probabilidad y la medición lo son. El uso de los conjuntos y los diagramas de Venn no es abundante en las ciencias de la conducta. Investigadores reconocidos a través de los años han empleado los conjuntos y diagramas de Venn en su trabajo de investigación. Piaget (1957), por ejemplo, usó el álgebra de conjuntos para tratar de explicar el pensamiento de los niños (véase también Piaget, García, Davidson y Easley, 1991). El trabajo clásico de Lewin (1935) en psicología Gestalt se apoya en conjuntos y diagramas de Venn para describir la interacción entre las personas y sus entornos, así como con ellos mismos. Más recientemente, Kubat (1993) aplicó conjuntos a estudios sobre aprendizaje de conceptos. Sheridan (1997) utilizó los diagramas de Venn para describir su marco conceptual para la intervención conjunta en cuestiones de conducta en psicología escolar. Sheridan establece que la identificación de problemas, el análisis de problemas, la implementación de planes y la evaluación del plan para la conducta de un niño puede explicarse como la intersección de los sistemas familiar, escolar y

[1] No queremos implicar que sea necesariamente deseable hacer el cambio, ni tampoco que toda la música moderna, aun toda la música de Ives, sea grandiosa o buena música. Tan sólo tratamos de mostrar la generalidad y aplicabilidad de las ideas de conjuntos y niveles de discurso.

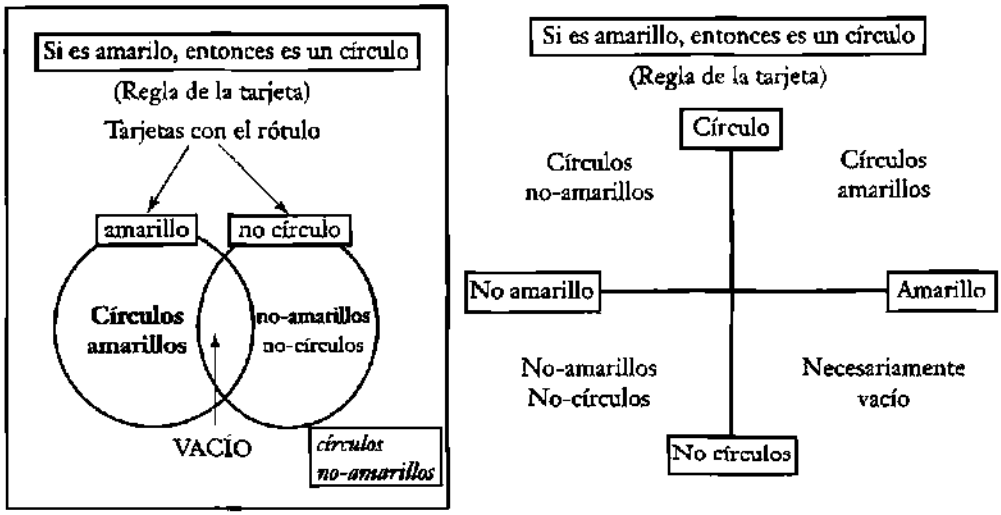
FIGURA 4.9



de apoyo del niño. Dayton (1976) emplea un diagrama de Venn para explicar cómo el individuo creativo confronta sus procesos mentales preconscientes y el mundo externo. Trites y Laprade (1983) los utilizan para ilustrar gráficamente un análisis de contingencias. Este análisis involucra un compuesto de puntuaciones factoriales en el estudio de la hiperactividad y el trastorno de conducta en niños (véase figura 4.9).

Bolman (1995) discute el papel y la necesidad del conocimiento de la ciencia de la conducta en la educación y práctica médicas. Para hacerlo, se apoya en diagramas de Venn para ilustrar lo que significa "fuerzas biopsicosociales", es decir, la intersección de las ciencias biológicas (anatomía, fisiología) con la psicología (sentimientos, identidad, metas) y la sociología y antropología (cultura, familia, ética). La combinación de estos tres factores constituye, en términos de Bolman, "la realidad clínica". Lane (1986) es un defensor de la enseñanza del razonamiento condicional (pensamiento lógico) a niños a través de diagramas de Venn y teoría de conjuntos. Lane condujo diversos estudios que comparaban diferentes materiales instruccionales para la enseñanza del pensamiento lógico. En cada uno de ellos

FIGURA 4.10



los diagramas de Venn (el enfoque de teoría de conjuntos) resultaron superiores a otros métodos en cuanto a desempeño inmediato, retención y transferencia del aprendizaje. La figura 4.10 presenta una muestra usada por Lane para comparar los diagramas de Venn con la tabla de lógica cartesiana. El concepto o la regla de la tarjeta bajo estudio es "si es amarillo, entonces es un círculo". Más adelante en este libro, se definirá medición a partir de una sola ecuación de teoría de conjuntos. Además, aclararemos los principios básicos del muestreo, del análisis y del diseño de investigación con base en conjuntos y teoría de conjuntos. Por desgracia, la mayor parte de los científicos sociales y educadores no están al tanto de la generalidad, poder y flexibilidad del pensamiento de conjuntos. Sin embargo, es posible predecir con certeza, que los investigadores en las ciencias sociales y educación encontrarán al pensamiento y teoría de conjuntos progresivamente más útiles para conceptualizar problemas teóricos de investigación.

RESUMEN DEL CAPÍTULO

1. Los conjuntos son útiles para entender los métodos de investigación. Son el fundamento del proceso descriptivo, lógico y del pensamiento y procesamiento analíticos. Constituyen la base del análisis numérico, categórico y estadístico.
2. Un conjunto es una colección bien definida de objetos o elementos. Dos formas de definir un conjunto son:
   a) Enlistar todos los miembros del conjunto.
   b) Aportar una regla para determinar si los objetos pertenecen o no al conjunto.
3. Los subconjuntos son partes del conjunto original. Si el conjunto entero constituye la población, entonces los subconjuntos son muestras de dicha población.
4. Las operaciones de conjuntos incluyen la intersección y la unión.
   a) La intersección está compuesta por los elementos comunes a dos o más conjuntos o subconjuntos; el símbolo es ∩.
   b) La unión es la combinación de los elementos no redundantes de dos o más conjuntos o subconjuntos; el símbolo es ∪.
5. El conjunto universal, U, se define como todos los elementos bajo consideración. Algunas veces se le llama población. El conjunto vacío, E, es el conjunto que no contiene miembros o elementos. También se llama conjunto *nulo*.
6. El conjunto de negación se simboliza por "~". Al colocar este símbolo antes de un conjunto se indica que contiene miembros del conjunto universal, U, que no están contenidos en el conjunto. Por ejemplo, $\sim A$ nos dice que este conjunto contiene todos los elementos en U que *no están en A*.
7. Una partición es la división de U en subconjuntos tales que cuando se combinan, U se restituye. Otra característica indispensable de una partición es que no haya elemento alguno de un subconjunto que se traslape con los elementos de otros subconjuntos.
8. Partición cruzada es la combinación de dos o más particiones diferentes. Las tablas de partición cruzada o tablas de contingencia son ejemplos de una partición cruzada que muestra la relación entre dos variables.

SUGERENCIAS DE ESTUDIO

1. Dibuje dos círculos traslapados, dentro de un rectángulo. Marque las siguientes partes: el conjunto universal U, los subconjuntos A y B, la intersección de A y B, y la unión de A y B.

a) Si estuviera trabajando en un problema de investigación con niños de 5o. grado, ¿qué parte del diagrama representaría a los niños de los que puede tomar su muestra?
b) ¿Qué representarían los conjuntos A y B?
c) ¿Qué podría significar la intersección de A y B?
d) ¿Cómo cambiaría el diagrama para representar un conjunto vacío? ¿Bajo qué condiciones un diagrama así tendría significado de investigación?

2. Considere la siguiente partición cruzada:

	Republicano (B_1)	Demócrata (B_2)
Masculino (A_1)		
Femenino (A_2)		

¿Cuál es el significado de los siguientes conjuntos?; es decir, ¿qué son, y cómo llamaríamos a cualquier objeto en estos conjuntos?
a) $(A_1 \cap B_1)$; $(A_2 \cap B_2)$
b) A_1; B_1
c) $(A_1 \cap B_1) \cup (A_1 \cap B_2) \cup (A_2 \cap B_1) \cup (A_2 \cap B_2)$
d) $(A_1 \cap B_1) \cup (A_2 \cap B_1)$

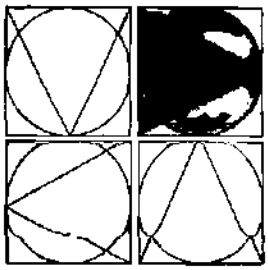
3. Genere una partición cruzada al utilizar las variables de estatus socioeconómico y preferencia de voto (demócrata y republicano). ¿Puede una muestra de individuos estadounidenses asignarse sin ambigüedad a las casillas de la partición cruzada? ¿Son exhaustivas las casillas? ¿No tienen traslapes? ¿Por qué son necesarias estas dos condiciones?
4. ¿Bajo qué condiciones puede ser verdadera la siguiente ecuación de conjuntos? [Observe que $n(A)$ representa el número de objetos en el conjunto A.]

$$n(A \cup B) = n(A) + n(B)$$

5. Suponga que un investigador en sociología quiere hacer un estudio sobre la influencia de la raza en el estatus ocupacional. ¿Cómo puede este investigador conceptualizar el problema en términos de conjuntos?
6. ¿Cómo están relacionados los conjuntos con las variables? ¿Podemos hablar sobre partición de variables? ¿Tiene sentido hablar de subconjuntos y variables? Explique.
7. Sea A = {Opus 101, Opus 106, Opus 109, Opus 110, Opus 111}, que es el conjunto de las cinco últimas sonatas para piano de Beethoven. Ésta es una definición de lista. He aquí una definición por regla:

$$A = \{a \mid a \text{ es una de las últimas cinco sonatas de Beethoven}\}$$

(El signo "|" se lee "dado"). ¿Bajo qué condiciones son mejores las definiciones de regla que las de lista?



CAPÍTULO 5

RELACIONES

Las relaciones son la esencia del conocimiento. Lo importante en ciencia no es el conocimiento de lo particular sino de las relaciones entre los fenómenos. Sabemos que las cosas son grandes sólo al compararlas con las más pequeñas. Así, establecemos las relaciones "mayor que" y "menor que". Los científicos educativos pueden "conocer" acerca del aprovechamiento sólo cuando estudian el aprovechamiento con relación al no aprovechamiento y a otras variables. Cuando saben que los niños de mayor inteligencia por lo general van bien en la escuela y que los niños de baja inteligencia con frecuencia lo hacen menos bien, ellos "conocen" una faceta importante del aprovechamiento. Cuando se percatan de que los niños de clase media tienden a desempeñarse mejor en la escuela que los de clase trabajadora, empiezan a comprender el "aprovechamiento". Conocen acerca de las relaciones que dan significado al concepto Aprovechamiento. Las relaciones entre Inteligencia y Aprovechamiento, entre Clase Social y Aprovechamiento y, de hecho, entre cualesquier variables constituyen el "meollo" básico de la ciencia.

La naturaleza relacional del conocimiento humano se ve con claridad aun cuando se analizan "hechos" en apariencia obvios. ¿Es dura una piedra? Para decir si este enunciado es cierto o falso primero debemos examinar conjuntos y subconjuntos de diferentes clases de piedras. Entonces, después de definir operacionalmente "duro" comparamos la "dureza" de piedras con otras "durezas". Los hechos más "simples" resultan, en el análisis, no

tan sencillos. Northrop (1947/1983, p. 317), al discutir sobre conceptos y hechos, afirma: "la única forma de obtener hechos puros, independientes de todo concepto y teoría, consiste tan sólo en observarlos para después permanecer perpetuamente mudo".

El diccionario nos dice que una *relación* es un vínculo, una conexión, un parentesco. Para la mayoría de las personas esta definición es suficientemente buena. Pero, ¿qué significan las palabras *vínculo*, *conexión* y *parentesco*? Otra vez, el diccionario dice que un *vínculo* es, una atadura, una fuerza que une; que una *conexión* es, entre otras cosas, una unión, una relación, una alianza. Pero una unión —un vínculo— ¿entre qué? Y ¿qué significa *unión*, *vínculo* y *fuerza que une*? Tales definiciones aunque intuitivamente útiles, son demasiado ambiguas para un uso científico.

Las relaciones como conjunto de pares ordenados

Las relaciones en ciencia siempre se dan entre clases o conjuntos de objetos. Uno no puede "conocer" la relación entre clase social y aprovechamiento escolar sólo al estudiar a un niño. "Conocer" la relación se logra sólo al abstraer la relación a partir de conjuntos de niños, o más precisamente, de conjuntos de características de niños. Permítanos tomar algunos ejemplos de relaciones y desarrollar de manera intuitiva un concepto de lo que es una relación.

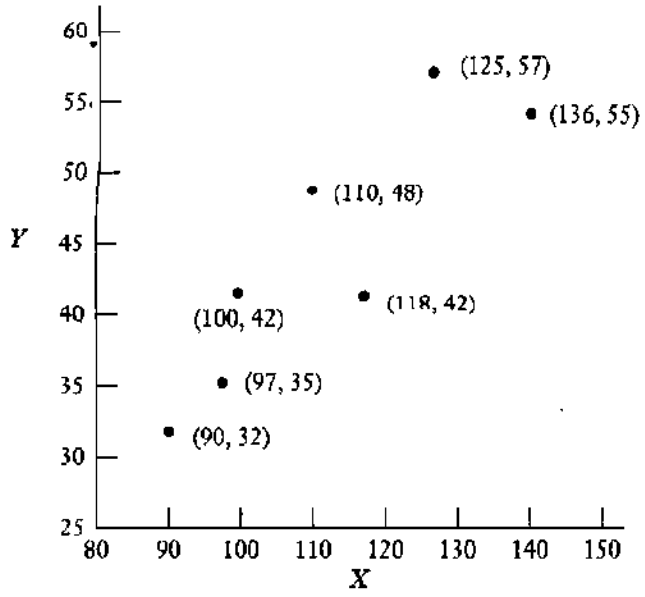
Sea A el conjunto de todos los papás, y B el de todos los hijos. Si pareamos a cada papá con su hijo (o hijos), tenemos la relación "papá-hijo". Podemos llamar a esta relación "crianza por parte del padre", aun cuando no se haya considerado a las hijas. De forma similar también podemos parear a los padres (elementos de A, en donde se considera a cada pareja de padres como un elemento individual) con sus hijos. Ésta constituirá la relación de "paternidad" o quizás, "familia". Sea A el conjunto de todos los esposos y B el conjunto de todas las esposas. El conjunto de parejas entonces define la relación "matrimonio". En otras palabras, un nuevo conjunto se ha formado, un conjunto de parejas en que se lista a los esposos siempre en primer término y a las esposas en segundo, y en donde cada esposo está aparejado sólo con su propia esposa.

Suponga que el conjunto A consiste de las puntuaciones de un grupo específico de niños en una prueba de inteligencia, y que B es la puntuación en una prueba de aprovechamiento. Si apareamos el coeficiente intelectual de cada niño con su puntuación de aprovechamiento, definimos una relación entre Inteligencia y Aprovechamiento. Observe que no podemos asignar con tanta facilidad un nombre como "paternidad" o "matrimonio" a está relación. Suponga que los conjuntos de puntuaciones son los siguientes:

Inteligencia	Aprovechamiento
136	55
125	57
118	42
110	48
100	42
97	35
90	32

Considere los dos conjuntos como un conjunto de pares, entonces este conjunto constituye una relación. Si graficamos ambos conjuntos de puntuaciones en los ejes X y Y, como se hizo en el capítulo 3 (figura 3.3), es más fácil "ver" la relación. Esto se hizo en la figura 5.1.

FIGURA 5.1

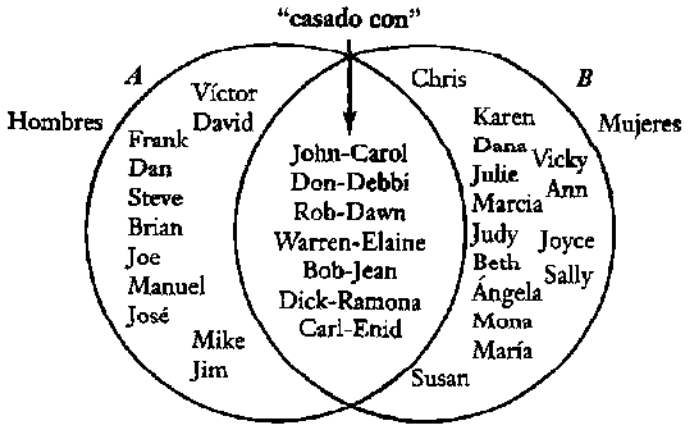


Cada punto se define por dos puntuaciones. Por ejemplo, el punto de la extrema derecha se define por (136, 55), y el de la extrema izquierda por (90, 32). Las gráficas como la figura 5.1 representan formas sucintas y muy útiles para expresar relaciones. Uno puede ver de inmediato, por ejemplo, que los valores más altos de X están acompañados por los valores más altos de Y, y los valores más bajos de X por los valores más bajos de Y. Como se verá en un capítulo posterior, también es posible trazar una línea a través de los puntos graficados de la figura 5.1, de la parte inferior izquierda a la parte superior derecha. (El lector debería tratar de hacerlo.) Esta línea, llamada *línea de regresión*, expresa la relación entre X y Y, entre inteligencia y aprovechamiento, pero también nos da, de una forma sucinta, considerablemente más información acerca de la relación: a saber: su dirección y magnitud.

Ahora estamos listos para definir "relación" de una manera formal: *una relación es un conjunto de pares ordenados*. Cualquier relación constituye un conjunto, un conjunto de cierta clase: un conjunto de pares ordenados. Un *par ordenado* está formado por dos objetos, o por un conjunto de dos elementos, en el que hay un orden fijo de aparición de los objetos. En realidad, hablamos de pares ordenados que significa (como antes se indicó) que los miembros de *cada par* siempre aparecen en un orden determinado. Si los miembros de los conjuntos A y B están pareados, entonces debemos especificar si los miembros de A o los de B aparecen primero en cada par. Si definimos la relación de matrimonio, por ejemplo, especificamos el conjunto de pares ordenados con, digamos, los esposos siempre en primer lugar en cada par. En otras palabras, el par (a, b) no equivale al par (b, a). Los pares ordenados están encerrados en paréntesis y marcamos: ( ), y un conjunto de pares ordenados se indica: $\{(a, k)\ (b, l), (c, m)\}$.

Por fortuna hemos dejado atrás la definición ambigua del diccionario. La definición de relaciones como conjuntos de pares ordenados, aunque puede parecer un poco extraña

FIGURA 5.2



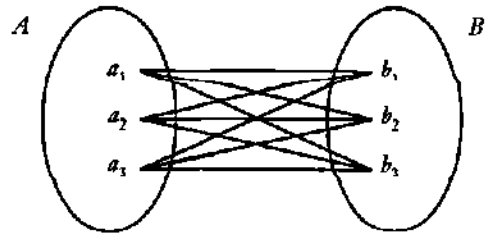
y aun curiosa al lector, no es ambigua y resulta general. Más aún, el científico, como el matemático, puede trabajar con ella.

Al discutir sobre relaciones, hay dos tipos especiales de conjuntos que juegan un papel importante. Uno se llama el *dominio* y el otro se llama la *imagen*. En lugar de definirlos de manera formal de inmediato, puede resultar más claro si consideramos un ejemplo: sea A el conjunto de todos los hombres y B el conjunto de todas las mujeres. Digamos que queremos definir la relación "casado con". Podemos hacerlo al formar la intersección apropiada de A y B (es decir, $A \cap B$) de manera que cada pareja ordenada en la intersección consistiera en las parejas casadas (mostrado en la figura 5.2). La intersección consiste en las parejas casadas. En este caso, el dominio en esta relación son los hombres que están casados. La imagen serían todas las mujeres que están casadas. Así, el dominio sería el conjunto de hombres que están en la intersección $A \cap B$ y la imagen sería el conjunto de mujeres que están en la intersección. El dominio es {John, Don, Rob, Warren, Bob, Dick y Carl}. La imagen es {Carol, Debbi, Dawn, Elaine, Jean, Ramona y Enid}. Observe que el dominio de la relación es siempre un subconjunto de A, y la imagen de la relación es un subconjunto de B.

De manera formal, si permitimos que RL represente la relación, que a sean los elementos del conjunto A, y b los elementos del conjunto B, entonces el dominio de RL es el conjunto de todas las cosas a tal que, para algunos b, el par ordenado (a, b) está en RL. La imagen (que también se llama contradominio) de la relación RL es el conjunto de todas las cosas b tal que, para algunas x, el par ordenado (a, b) está en RL.

Definir el dominio y la imagen en una relación es importante porque juegan un papel clave al definir una *función*. Hays (1994) considera a la función como uno de los conceptos más importantes en matemáticas y ciencia. Las funciones y las relaciones son muy similares. Se puede considerar que una función es una clase especial de relación. Una relación es una función cuando cada elemento del dominio está pareado con un miembro y sólo con uno de la imagen. La mayoría de las personas concibe a la función en términos numéricos, pero esto no es necesariamente así. En la sociedad estadounidense, por ejemplo, la relación de ser esposo es una función, en tanto que un hombre tiene a lo más una esposa en cualquier momento dado. Sin embargo, la relación de ser una madre no es una función ya que una persona puede ser madre de varios niños. Si lo vemos con cuidado, la relación de ser la hija de una madre sí es una función, en tanto que esa niña puede tener sólo una

FIGURA 5.3



madre biológica. Así, una función es un conjunto de pares ordenados, en donde no hay dos pares distintos que posean los mismos elementos.

Determinación de relaciones en la investigación

Aunque hemos evitado la ambigüedad con nuestra definición de relaciones, no hemos aclarado en especial el problema práctico de "determinar" relaciones. Existe otra forma de definir una relación que nos puede ayudar. Sean A y B conjuntos. Si pareamos de manera individual cada miembro de A con cada miembro de B, obtendremos *todos los pares posibles* entre ambos conjuntos, lo que se denomina el *producto cartesiano* de los dos conjuntos y se enuncia $A \times B$. Una relación se define como un subconjunto de $A \times B$; es decir, cualquier subconjunto de pares ordenados tomados de $A \times B$ constituye una relación. (véase Kershner y Wilcox, 1974, para un excelente análisis de relaciones).

Para ilustrar esta idea de manera sencilla, sea el conjunto $A = \{a_1, a_2, a_3\}$ y el conjunto $B = \{b_1, b_2, b_3\}$.[1] Entonces, el producto cartesiano, $A \times B$, puede diagramarse como se muestra en la figura 5.3. Esto es, generamos nueve pares ordenados: (a_1, b_1), (a_1, b_2), ..., (a_3, b_3). Con conjuntos grandes, por supuesto, habrá muchos pares, de hecho tendremos mn pares, donde m y n son la cantidad de elementos en A y B, respectivamente.

Esto no es muy interesante —al menos en este contexto—. ¿Qué hacemos para determinar o "descubrir" una relación? De manera empírica determinamos qué elementos de A "van con" qué elementos de B, de acuerdo con algún criterio. Es obvio que hay muchos subconjuntos de pares de $A \times B$, la mayoría de los cuales no "tienen sentido" o no nos interesan. Kershner y Wilcox (1974) afirman que una relación es "un método para distinguir ciertos pares ordenados de otros; es un esquema para señalar determinados pares de todos los demás". De acuerdo con esta forma de concebir las relaciones, la relación de "matrimonio" es un método o procedimiento para distinguir las parejas casadas de todos los posibles pares de hombres y mujeres. De esta forma podemos incluso considerar a la religión como una relación. Sea $A = \{a_1, a_2, \ldots, a_n\}$ el conjunto de todas las personas en los Estados Unidos y sea $B = \{$Católica, Protestante, Judía, etcétera$\}$ el conjunto de religiones. Si ordenamos los pares, en este caso cada persona con una religión, entonces tenemos la "relación" de religión o, para ser más precisos, la "afiliación religiosa". Para evitar que el estudiante esté confundido por la extraña sensación de definir una relación como un subconjunto de $A \times B$, diremos otra vez que es natural que muchos de los posibles subconjuntos de pares ordenados $A \times B$ no tengan sentido. Quizá lo más importante

[1] Los subíndices sólo etiquetan y distinguen a los miembros individuales de los conjuntos. No implican orden alguno. Observe también que no es necesario que haya igual número de miembros en ambos conjuntos.

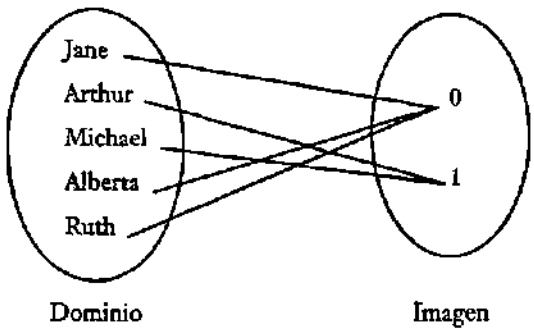
sea que nuestra definición de relación no presenta ambigüedad y es general por completo. No importa qué conjuntos de pares ordenados elijamos, constituyen una relación. Nos corresponde decidir cuáles conjuntos tienen sentido científico según el dictado de los problemas para los que buscamos respuestas y cuáles no.

El lector puede preguntarse por qué nos tomamos tantas molestias en definir las relaciones. La respuesta es simple: casi toda la ciencia busca y estudia relaciones. Literalmente no existe forma empírica para "conocer" nada, excepto a través de sus relaciones con otras cosas, como lo indicamos antes. Si, como en el caso de Behling y Williams (1991), nos interesamos en la percepción de la inteligencia y las expectativas del aprovechamiento escolar, es necesario que relacionemos la percepción y la expectativa con otras variables. Para explicar un fenómeno como la percepción de la inteligencia debemos "descubrir" sus determinantes —las relaciones que tiene con otras variables pertinentes—. Behling y Williams "explicaron" las percepciones de los maestros sobre la inteligencia y las expectativas del aprovechamiento escolar hacia los estudiantes al relacionarlos con el tipo y estilo de vestimenta de los estudiantes de educación media superior. Encontraron que el estilo de vestimenta influye en las percepciones tanto de los maestros como de los compañeros. Es evidente que, si las relaciones son fundamentales en la ciencia, debemos saber con claridad qué son, al igual que cómo se estudian. Se ha descuidado la definición de "relación" en la investigación del comportamiento. Parece que se asume que es un concepto cuyo significado todos conocen. También se le confunde con "relación" que es una conexión de alguna clase entre la gente, o entre la gente y los grupos como una relación madre-hijo. *No es lo mismo que una relación.*

Reglas de correspondencia y mapeo

Cualquier objeto —gente, números, resultados en juegos de azar, puntos en el espacio, símbolos, etcétera— pueden ser miembros de conjuntos y pueden relacionarse como pares ordenados. Se dice que los miembros de un conjunto son *mapeados* en los miembros de otro conjunto al usar una regla de correspondencia. Una *regla de correspondencia* es una receta o una fórmula que nos dice cómo mapear los objetos de un conjunto en los objetos de otro conjunto. Nos indica en pocas palabras, cómo debe realizarse la correspondencia entre los miembros de los conjuntos. Estudie la figura 5.4, que muestra la relación entre los nombres de cinco individuos y los símbolos 1 y 0, que representan masculino (1) y femenino (0). Tenemos aquí un mapeo de sexo (1 y 0) en los nombres. Esto es por supuesto, una relación donde cada nombre tiene asignado un 1 o un 0, masculino o femenino.

FIGURA 5.4



En una relación, los dos conjuntos cuyos "objetos" se relacionan se llaman *el dominio* y la *imagen* o *contradominio*, o D y C. D es el conjunto de los primeros elementos y C el conjunto de los segundos elementos. En la figura 5.4, le asignamos 1 a masculino y 0 a femenino. A cada miembro del dominio se asigna el miembro apropiado de la imagen. D = {Jane, Arthur, Michael, Alberta, Ruth}, y C = {0, 1}. La regla de correspondencia, dice: si el objeto de D es femenino asigne un "0", si es masculino un "1".

En otras palabras, los objetos, especialmente los números, se asignan a otros objetos —personas, lugares, números, etcétera— de acuerdo a la regla. El proceso es muy variado en sus aplicaciones pero simple en su concepción. En lugar de pensar en todas las diferentes formas de expresar las relaciones por separado, caemos en la cuenta de que todos son conjuntos de pares ordenados y que los objetos de un conjunto simplemente se mapean en los objetos del otro conjunto. Todas las variadas formas de expresar relaciones —como mapeos, correspondencias, ecuaciones, conjuntos de puntos, tablas, o índices estadísticos— pueden reducirse a conjuntos de pares ordenados.

Algunas formas de estudiar relaciones

Podemos expresar las relaciones de varias formas. En el análisis previo, se ilustraron algunas. Una de ellas consiste simplemente en listar y parear los miembros de los conjuntos, como en las figuras 5.3 y 5.4. De hecho, este método no se usa con frecuencia en la literatura de investigación. Ahora se examinarán algunos procedimientos más útiles.

Gráficos

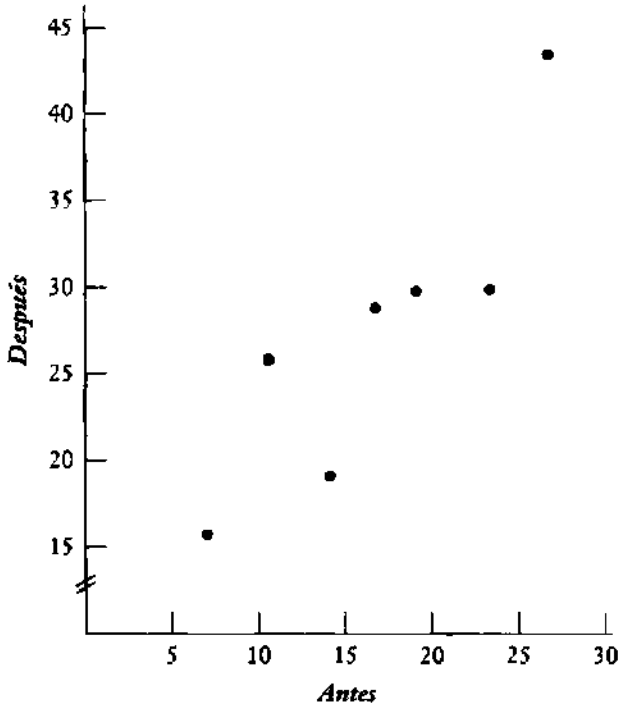
Un gráfico es un dibujo en el que los dos miembros de cada par ordenado de una relación se trazan en dos ejes, X y Y (o cualquier otra designación apropiada). La figura 5.1 es un gráfico de pares ordenados de las puntuaciones ficticias de inteligencia y aprovechamiento que se dieron antes. Podemos ver en la gráfica que los pares ordenados tienden a "ir juntos": los valores altos de Y van con los valores altos de X, y los valores bajos de Y van con los valores bajos de X.

Un conjunto más interesante de pares ordenados está graficado en la figura 5.5. Los números usados para este gráfico fueron tomados de un estudio fascinante de Miller y DiCara (1968), en el que se entrenó a siete ratas para secretar orina. (Dado que la secreción de orina es una función autónoma está normalmente más allá de control y por lo tanto, del entrenamiento y aprendizaje.) El "antes" o eje de las X de la gráfica indica valores de secreción de orina antes del entrenamiento; el "después" o eje de las Y indica valores después del entrenamiento. Utilizaremos los mismos datos en otro contexto más adelante en el libro (y describiremos el estudio más detenidamente), por ahora no habrá más detalles. La relación entre los dos conjuntos de valores de secreción de orina es pronunciada. Otra vez, valores elevados antes del entrenamiento se acompañan por valores altos después del entrenamiento, y sucede algo similar con los valores bajos. El gráfico y la relación que expresa refleja las diferencias individuales en la secreción de orina. El significado completo de este enunciado será claro cuando más tarde describamos el análisis estadístico de estos datos.

Tablas

Quizás la forma más común de presentar datos para mostrar relaciones es a través de tablas. Las variables de las relaciones presentadas se señalan por lo general en la parte alta

FIGURA 5.5



	Antes	Después
1	23	30
2	14	19
3	16	29
4	18	30
5	7	16
6	26	44
7	12	26

y a los lados de la tabla, mientras que los datos están contenidos en su interior. Los datos estadísticos son casi siempre medias, frecuencias y porcentajes. Considere la tabla 5.1, que es una presentación resumida de los datos de frecuencias presentados por Freedman, Wallington y Bless (1967). Estos investigadores sometieron a prueba la noción de que el acatamiento se relaciona con la culpa: a mayor sentimiento de culpa, mayor será el acatamiento a las peticiones. Los experimentadores indujeron a la mitad de los sujetos a mentir, y supusieron que hacerlo produciría culpa (al parecer así fue). La variable Culpa o Mentira está localizada en la parte superior de la tabla. Ésta es la variable independiente, por supuesto. La variable dependiente fue el acatamiento a las peticiones hechas. Esta variable está etiquetada a un lado de la tabla. Los datos en las casillas de la tabla son frecuencias; esto es, el número de sujetos que cayeron dentro de los subconjuntos o subcategorías. De los 31 sujetos inducidos a mentir, 20 acataron las demandas del

TABLA 5.1 *Resultados en frecuencias del experimento para estudiar la relación entre culpa y acatamiento (estudio de Freedman, Wallington y Bless).*

	Mentir (Culpa)	Sin mentir (No culpa)
Acatamiento	20	11
No acatamiento	11	20
	31	31

experimentador y 11 no. De los 31 sujetos que no fueron inducidos a mentir, 11 acataron y 20 no lo hicieron. Los datos son consistentes con la hipótesis. En un capítulo posterior se estudiará en detalle cómo analizar e interpretar datos de frecuencia y tablas de esta naturaleza.

El punto esencial de la tabla 5.1 es que la relación y la evidencia de la naturaleza de la relación se expresan en ella. En este caso los datos se presentan en forma de frecuencias. (Una frecuencia es el número de miembros de conjuntos y subconjuntos. Un porcentaje es una razón o proporción por ciento. Se calcula al multiplicar 100 veces la razón de un subconjunto a un conjunto, o de un subconjunto a otro subconjunto.) La tabla en sí constituye una partición cruzada, con frecuencia llamada tabulación cruzada o tabla de contingencia en la que una variable de la relación se entrecruza contra otra variable de la relación. Los nombres de ambas variables aparecen en la parte superior y al lado de la tabla, como se indicó antes. La dirección y magnitud de la relación en sí misma se expresa por los tamaños relativos de las frecuencias en las casillas de la tabla. En la tabla 5.1 muchos más sujetos (20 de 31) inducidos a mentir acataron la petición que los sujetos no inducidos a mentir (11 de 31).

Se presenta un ejemplo más complicado en la tabla 5.2. Se trata de una presentación resumida de los datos de frecuencia de un estudio de Mays y Arrington (1984). Estos investigadores sometieron a prueba la noción de que la violación de la territorialidad o límites espaciales está relacionada con características demográficas. Las personas con características de un bajo estatus (raza o género) tienden a que se les viole su espacio con mayor frecuencia. Los experimentadores crearon 10 condiciones para representar 10 configuraciones diádicas. Estas diadas se generaron al cruzar dos niveles de raza (caucásico y afroamericano) con dos niveles de sexo (femenino y masculino); por ejemplo, "afroamericano femenino con caucásico masculino". Los confederados con dichas especificaciones demográficas para cada una de 10 condiciones planteadas se mantuvieron a una distancia confortable para conversar y sostuvieron una discusión casual. Dos observadores ocultos registraron la ruta que siguieron unas 210 personas que ingresaban en cada condición. Anotaron las tendencias para atravesar (y así penetrar los límites de las diadas) o pasar alrededor de ellas. También registraron el grupo étnico y el género de los que ingresaban. La variable independiente es la configuración diádica. La variable dependiente es la penetración en los límites de la diada o la no penetración (pasar alrededor de ella). Aunque Mays y Arrington consideraron muchas variables en el estudio, por ejemplo, aquí sólo usaremos la combinación diádica de sexo (masculino-masculino, femenino-femenino, femenino-masculino). Las combinaciones se encuentran etiquetadas en la parte superior de la tabla 5.2 y la variable dependiente está al lado. La tabla muestra tanto las frecuencias como los porcentajes.

Los porcentajes totales para las combinaciones masculino-masculino y femenino-femenino fueron muy similares (31% *vs.* 29%). También, los porcentajes de la conducta de rodear para las diadas masculino y femenino fueron muy parecidos (30% *vs.* 28%). Sin

TABLA 5.2 *Resultados en frecuencias y porcentajes del experimento para estudiar la relación entre diadas sexuales y violación del espacio (estudio de Mays y Arrington)*

	Masculino-Masculino	Femenino-Femenino	Femenino-Masculino	
Alrededor	650 (30)	600 (28)	898 (42)	2148 (86)
A través de	106 (33)	119 (37)	98 (30)	323 (14)
	756 (31)	719 (29)	996 (40)	2471

embargo, la tabla muestra que hay una tendencia de las diadas femenino-masculino para rodear (42%). Uno puede especular, a partir de ello, que hay un mayor respeto por el espacio compartido por parte de combinaciones de sexo diferente que del mismo. Mays y Arrington dan crédito a esta noción al señalar los porcentajes que aparecen en la violación o en los datos de "a través de" (segundo reglón de la tabla 5.2). En términos de porcentajes, la porción de espacio de las diadas femeninas se viola con más frecuencia que las masculinas (37% *vs.* 33%). Para las diadas con diferente combinación de sexo (femenino-masculino), el espacio se invade menos que para las otras (30%). En un capítulo posterior se estudiará en detalle cómo analizar e interpretar los datos de frecuencia (porcentaje) y las tablas de esta clase.

La tabla 5.2 muestra la naturaleza del formato de la tabla de relación. En este caso, los datos están en forma de frecuencias y porcentajes. Usar los porcentajes de la tabla 5.2 para un análisis visual simple, resulta más fácil que utilizar las frecuencias puras. Con los porcentajes, en la tabla 5.2, el valor máximo es 100 y el mínimo es 0. Hay más invasores de la diada femenino-femenino (119 de 323 o 37%) que en la diada masculino-masculino (106 de 323 o 33%).

Una clase diferente de tabla presenta medias, promedios aritméticos, en el cuerpo del cuadro. Las medias expresan la variable dependiente. Si hay sólo una variable independiente, sus categorías estarán etiquetadas en la parte superior de la tabla. Si hay dos o más variables independientes, sus categorías pueden presentarse de varias formas en la parte superior y a los lados, como se verá en capítulos posteriores. Se da un ejemplo en la tabla 5.3, que constituye la forma más simple que una tabla puede asumir. Hyatt y Tingstrom (1993) estudiaron el efecto del uso de jerga conductual en la percepción de los maestros hacia dos intervenciones conductuales: reforzamiento y castigo. Los hallazgos en investigaciones previas en el efecto de la jerga de los maestros no fueron concluyentes. En este estudio, se presentó a los maestros con un estudiante hipotético con un problema conductual. Luego se les dio la descripción de dos tipos de intervenciones. La descripción contenía o bien jerga conductual tal como "condicionado de manera operante y comportamiento apropiado incompatible", o bien, una descripción sin jerga con palabras tales como "se le premió por sentarse correctamente". Los maestros que recibieron la jerga conductual pueden considerarse como del grupo experimental, mientras que los que recibieron instrucciones sin jerga constituyeron el grupo control. Todas las percepciones de los maestros fueron medidas por medio del Inventario de Evaluación del Tratamiento (también referido como TEI), que permite a los maestros evaluar las intervenciones en términos de cómo perciben la aceptabilidad, adecuación, justicia y efectividad. Las puntuaciones variaron de entre 15 a 105. Las puntuaciones altas indicaban una mayor aceptabilidad. Como puede verse en la tabla 5.3, la media del grupo experimental es mayor que la del grupo control. ¿Es "grande" o "pequeña" la diferencia entre las medias? Más adelante se verá cómo evaluar el tamaño y el significado de tales diferencias. Por ahora, sólo nos interesa por qué la tabla expresa una relación.

En tablas de esta clase siempre se expresa o implica una relación. Las que son tan simples como ésta, rara vez se usan en la literatura. Basta sólo mencionar ambas medias en

TABLA 5.3 *Medias de los grupos con y sin jerga (estudio de Hyatt y Tingstrom)*[a]

Experimental (jerga)	Control (sin jerga)
79.38	73.68

[a] Las medias se calcularon a partir del Inventario de Evaluación del Tratamiento.

el texto de un informe. Más aún, pueden compararse más de dos medias. Sin embargo, el principio es el mismo; las medias "expresan" la variable dependiente y las diferencias entre ellas, el efecto supuesto de la variable independiente. En este caso hay dos variables relacionadas: la jerga y la percepción. La rúbrica "experimental y control" expresa la jerga recibida por el grupo experimental pero no por el control. Ésta es la variable independiente. Las dos medias en el cuadro expresan la percepción del maestro con relación al método de intervención, medida por el TEI: ésta es la variable dependiente. Si las medias difieren lo suficiente, puede asumirse que la jerga conductual tuvo un efecto en la percepción de los maestros o su aceptabilidad.

Las tablas de medias son en extremo importantes en la investigación del comportamiento, en especial en investigación experimental. Puede haber dos, tres o más variables independientes, y pueden expresar los efectos individuales y combinados de estas variables en una variable dependiente, o incluso en dos o más variables dependientes. El punto central es que siempre se estudian las relaciones, aun cuando no siempre es fácil conceptualizarlas y establecerlas.

Gráficos y correlación

Aunque antes examinamos de forma breve las relaciones y gráficas, puede ser provechoso ahondar en este tópico. Suponga que tenemos dos conjuntos de puntuaciones de los mismos individuos en dos pruebas, X y Y:

X	Y
1	1
2	1
2	2
3	3

Los dos conjuntos forman un conjunto ordenado de pares. Este conjunto constituye, por supuesto, una relación. También puede escribirse, usando R para denotar relación, $R = \{(1,1), (2,1), (2,2), (3,3)\}$. Se diagrama en la gráfica de la figura 5.6.

Con frecuencia podemos darnos una idea aproximada de la dirección y grado de la relación al inspeccionar la lista de pares ordenados, pero es un método impreciso. Las

FIGURA 5.6

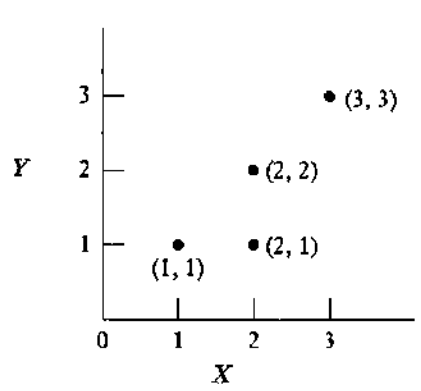
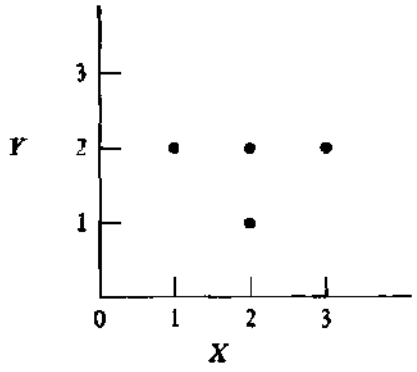


FIGURA 5.7



gráficas, como la de las figuras 5.1 y 5.6, nos dicen más. Es más fácil "ver" que los valores de Y "se mueven" con los valores de X. Los valores más altos de X acompañan a los valores más altos de Y, y viceversa. En este caso, la relación —o correlación, como también se le llama— es positiva. (Si tuviéramos la ecuación $R = \{(1,3), (2,1), (2,2), (3,1)\}$, la relación sería negativa. El lector debe graficar estos valores y observar su dirección y significado). Si la ecuación fuera $R = \{(1,2), (2,1), (2,2), (3,2)\}$, la relación sería nula o cero; esto se grafica en la figura 5.7. Puede verse que los valores de Y no "se mueven" con los valores de X de ninguna forma sistemática. Esto no significa que "no haya" relación. Siempre existe una relación —por definición— en tanto que existe un conjunto de pares ordenados. Sin embargo, es común que se diga que "no hay" relación. Resulta más preciso decir que la relación es nula o cero.

Los científicos sociales por lo general calculan índices de relación, con frecuencia llamados coeficientes de correlación, entre conjuntos de pares ordenados para obtener estimaciones más precisas de la dirección y grado de las relaciones. Si calculamos uno de estos índices, el coeficiente de correlación producto-momento, o r, para los pares ordenados de la figura 5.6, obtenemos $r = .85$. Para los pares de $R = \{(1,3), (2,1), (2,2), (3,1)\}$ dijimos que la relación era negativa, $r = -.85$. Para los pares de la figura 5.7, el conjunto de pares mostró una relación nula o cero, $r = 0$.[2]

El coeficiente producto-momento y otros coeficientes de correlación, están basados en la variación concomitante de los miembros de conjuntos de pares ordenados. Si *covarían* —varían juntos altos valores con altos valores, valores medios con valores medios, y valores bajos con valores bajos, o valores altos con valores bajos, etcétera— se dice que hay una relación positiva o negativa, en su caso. Si no covarían, se dice que "no hay" relación. Los índices más útiles varían de +1.00 a través de 0 hasta –1.00. Un +1.00 indica una relación positiva perfecta, –1.00 indica una relación negativa perfecta, y un 0 indica relación no discernible (o relación cero). Algunos índices van sólo de 0 a +1.00. Otros índices pueden tomar valores distintos.

La mayoría de los coeficientes de relación nos dicen qué tan similares son los órdenes de posición de los dos conjuntos de medidas. La tabla 5.4 presenta tres ejemplos para ilustrar estos órdenes de posición o formas de "moverse" juntos. Se presentan los coefi-

[2] Los métodos para calcular estas r y otros coeficientes de correlación se discuten en los textos de estadística, que incluyen un análisis de mayor profundidad sobre la interpretación de coeficientes de correlación de la que es posible en este libro.

TABLA 5.4 *Tres conjuntos de pares ordenados con diferentes direcciones y grados de correlación.*

(*I*) $r = 1.00$		(*II*) $r = -1.00$		(*III*) $r = 0$	
X	*Y*	*X*	*Y*	*X*	*Y*
1	1	1	5	1	2
2	2	2	4	2	5
3	3	3	3	3	3
4	4	4	2	4	1
5	5	5	1	5	4

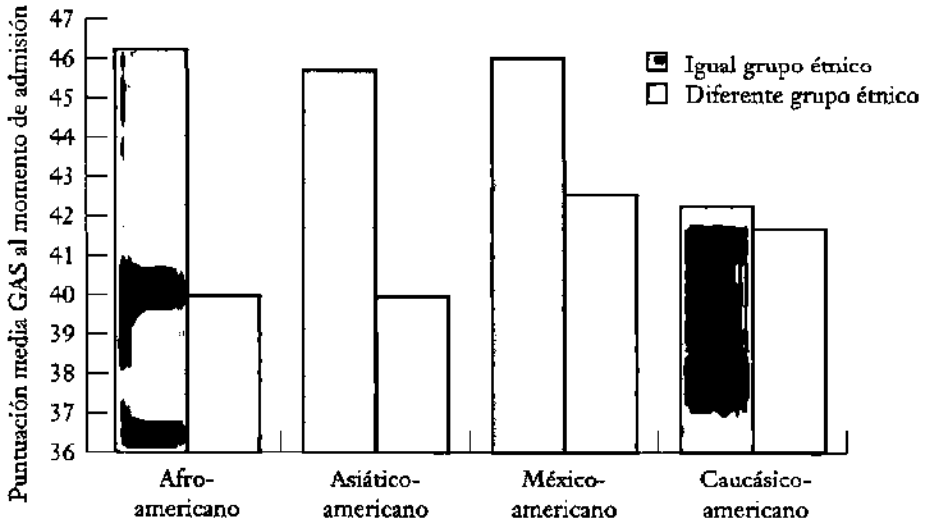
cientes de correlación con cada uno de los conjuntos de pares ordenados. *I* es obvio: el orden de las puntuaciones *X* y *Y* de *I* se mueven perfectamente en conjunto. Así lo hacen las puntuaciones *X* y *Y* de *II*, pero en la dirección opuesta. En *III*, no hay relación discernible entre la posición del orden. En *I* y *II*, uno puede predecir a la perfección el valor de *Y* dado el valor de *X*, pero en *III* uno no puede predecir los valores de *Y* a partir de *X*. Rara vez los coeficientes de correlación son 1.00 o 0; por lo común toman valores intermedios.

Ejemplos de investigación

Para ilustrar nuestra abstracta discusión sobre las relaciones, veamos dos ejemplos interesantes de relaciones y correlación. Russell, Fujino, Sue, Cheung y Snowden (1996) examinaron los efectos de la etnicidad en las diadas terapeuta-cliente en la evaluación del funcionamiento mental. Estos investigadores usaron un gran conjunto de datos de clientes adultos atendidos en los servicios de consulta externa de una importante clínica metropolitana de salud mental. De estos datos, los investigadores extrajeron cuatro grupos étnicos para el estudio: asiático-americanos, afro-americanos, méxico-americanos, y caucásico-americanos. La Escala de Evaluación Global (GAS por sus siglas en inglés) obtenida al momento de la admisión, se usó como medida del funcionamiento mental. El terapeuta del caso asignó la puntuación del GAS al cliente. Las puntuaciones altas indicaban un buen funcionamiento general, mientras que las bajas señalaban una severa disfunción.

Russell y sus colaboradores examinaron las puntuaciones del GAS para aquellos clientes cuyo origen étnico era el mismo del terapeuta (por ejemplo, terapeuta asiático-americano con cliente asiático-americano; terapeuta afro-americano con cliente afro-americano, etcétera) contra las puntuaciones del GAS de aquellos clientes que difirieron en cuanto a origen étnico con relación a su terapeuta. La figura 5.8 muestra la relación entre unos y otros, y sus puntajes GAS. Observe que aquí la relación se presenta en forma de una gráfica diferente a las que hemos visto antes. Este formato se denomina *histograma* o *gráfico de barras*. El gráfico muestra consistentemente que las puntuaciones GAS fueron más altas (mejor salud mental) cuando los terapeutas fueron del mismo origen étnico que los clientes, que cuando no lo fueron. Esto significa que el terapeuta percibió al cliente con un mayor nivel del funcionamiento en cuanto a salud mental cuando el cliente tenía el mismo origen étnico, que cuando el cliente era de un grupo étnico diferente. También para las minorías étnicas, cuando había coincidencia de terapeuta-cliente en este sentido, se presentaba una mayor puntuación GAS que en el caso de caucásicos. Sin embargo, para los clientes en cuyos casos no había correspondencia étnica, la puntuación GAS fue la más alta para los méxico-americanos y la más baja para los afro-americanos y asiático-americanos. La mayor discrepancia entre las relaciones con y sin coincidencia étnica en la puntuación de GAS fue en los grupos minoritarios.

FIGURA 5.8



Nuestro segundo ejemplo no es cuantitativo, aunque implica cantidad y no sería difícil cuantificar las variables. Hardy (1974) estudió, entre otras cosas, la relación entre afiliación religiosa y productividad doctoral. ¿Qué grupo religioso produce doctorados más sobresalientes y cuál los menos? (Hardy estudiaba en realidad valores y su influencia en la escolaridad.) Los resultados se presentan en la tabla 5.5. Requieren comentario: es aparente que la relación es fuerte; mientras más liberal es un grupo religioso, mayor producción de grados doctorales. Los pares ordenados de grupos religiosos y su evaluación de productividad se ven fácilmente.

Nuestro último ejemplo es un estudio de Little (1997). Constituye una variante del estudio de Hardy presentado antes. A diferencia de esa investigación, el estudio de Little incluye un alto nivel de cuantificación: estudió la relación entre las universidades que ofrecen grados y la productividad académica en el campo de la psicología escolar. La pregunta fue: ¿qué posgrado universitario produce a los graduados más sobresalientes? La respuesta es importante porque aportará información más allá de los resultados de la investigación previa, que se basó sobre todo en la afiliación institucional de los autores sin

TABLA 5.5 *La relación entre afiliación religiosa y resultados de doctores sobresalientes en los Estados Unidos (estudio de Hardy).*

Tipo de religión	Evaluación de productividad
Liberal, protestante secularizado y judíos	Alta productividad
Liberal moderado, disidente, protestante antitradicional	Productividad arriba del promedio
Protestante tradicional	Productividad moderada
Fundamentalistas y protestantes conservadores	Baja productividad
Católicos	Productividad muy baja

TABLA 5.6 *La relación entre la educación a nivel posgrado y el número de graduados en psicología escolar; de 1987 a 1995 (estudio de little).*

Orden	Universidad	Número de graduados	Total ponderado
1	Georgia	19	65.98
2	Indiana	10	52.48
3	Minnesota	13	40.88
4	Texas	12	38.61
5	Wisconsin	11	27.92
6	Columbia	10	27.60
7	California, Berkeley	7	27.24
8	South Carolina	4	22.39
9	Oregon	5	21.48
10	Ball State	7	20.32
11	Ohio State	7	19.56
12	Kent State	5	19.36
13	Nebraska	8	18.31
14	Arizona State	2	16.28
15	Utah	5	15.90
16	Temple	5	15.88
17	Indiana State	4	15.61
18	Illinois	2	13.43
19	Southern Mississipi	7	13.36
20	Connecticut	4	12.75
21	Michigan State	5	12.38
22	Pittssburgh	1	11.97
23	Cincinnati	5	11.91
24	Pennsylvania	3	11.03
25	Penn State	4	10.09

proporcionar información acerca de dónde fueron educados. El estudio Little provee esa información al presentar los datos sobre el sitio en que los autores recibieron su grado terminal. Little establece que esta medida constituye un mejor indicador de la calidad de los programas de posgrado en psicología escolar. Una reproducción parcial de los resultados totales se muestra en la tabla 5.6. Little muestra que la mayoría de los programas de grado en Estados Unidos se concentra en particular en las regiones del medio oeste, sureste y en la costa este. Entre los datos de Little está un conjunto de pares ordenados que relacionan universidad y productividad. Little, en este estudio, encontró un buen número de discrepancias entre sus hallazgos y los publicados por *US News & World Report* (1995) sobre los mejores programas de Estados Unidos en esta disciplina. Los resultados de Little se basaron en datos empíricos reunidos en las seis revistas más importantes de investigación en psicología escolar. *US News & World Report* basó sus ordenamientos en la reputación de la universidad.

Relaciones multivariadas y regresión

En nuestro análisis sobre relaciones pudimos haber dado la impresión de que los científicos e investigadores están siempre preocupados por las relaciones entre dos variables. Cuando, por ejemplo, hablamos acerca de las relaciones entre coincidencia de origen étnico y salud mental, jerga y percepción, institución de posgrado y producción de trabajo académico, es

posible que hayamos generado la idea errónea de que los científicos se preocupan por estudiar sólo relaciones de dos variables. Esto no es así. De hecho, ha habido un gran número de investigación de dos variables, pero en las ciencias del comportamiento esto ha cambiado de forma dramática. La preocupación de los investigadores de la conducta en la actualidad tiene más que ver con las relaciones múltiples. Mientras que los investigadores modernos saben que la relación entre inteligencia y aprovechamiento es sustancial y positiva, también asumen que existen muchos determinantes para una y otra variable. Saben, por ejemplo, que la clase social tiene una influencia decisiva en ambas. También creen, aunque hay evidencia contradictoria, que la autoestima afecta a los dos. Más aún, los metodólogos han desarrollado poderosos enfoques y métodos analíticos para manejar lo que llamaremos problemas multivariados. Revisemos de manera breve la lógica y sustancia de estos problemas.

Algo de lógica de la investigación multivariada

La estructura oculta de nuestro argumento hasta ahora ha sido resumida por la expresión "si p, entonces q": "si inteligencia, entonces aprovechamiento"; "si estatus bajo, entonces violación del espacio"; "si este estilo de vestimenta, entonces esta percepción de inteligencia". Representan, desde luego, relaciones implicadas. Pero van más allá: también implican una dirección: de las variables independientes a las variables dependientes. Pueden concebirse como "si p, entonces q", que en lógica se denomina enunciado condicional. Es posible conceptualizar a la mayoría de los problemas de investigación y estudiar la estructura de los argumentos científicos al usar enunciados condicionales y otros relacionados (Kerlinger, 1969). Sin embargo, las relaciones de la investigación del comportamiento son más complejas que un simple enunciado "si p, entonces q". Es más probable que los investigadores contemporáneos digan algo como "si p, entonces q bajo las condiciones r y t". Este enunciado condicional puede escribirse como: $p \rightarrow q \mid r, t$, que se lee como en la oración precedente (" $\mid$ " significa "bajo las condiciones" o "dado"). De forma más simple podemos escribir: $(p_1, p_2, p_3) \rightarrow q$ que significa "si p_1 y p_2 y p_3, entonces q". De manera más concreta, esto significa que las variables p_1 y p_2 y p_3 influyen en la variable q en ciertas formas. Podemos decir, por ejemplo, que la inteligencia, clase social y autoestima afectan el rendimiento escolar de tal y cual formas.

La forma más simple de demostrar las relaciones en forma gráfica es a través de los diagramas de ruta. Un diagrama de ruta para el enunciado anterior se puede observar en la figura 5.9. En él, donde usamos x_1, x_2 y x_3 para las variables independientes y y para la

FIGURA 5.9

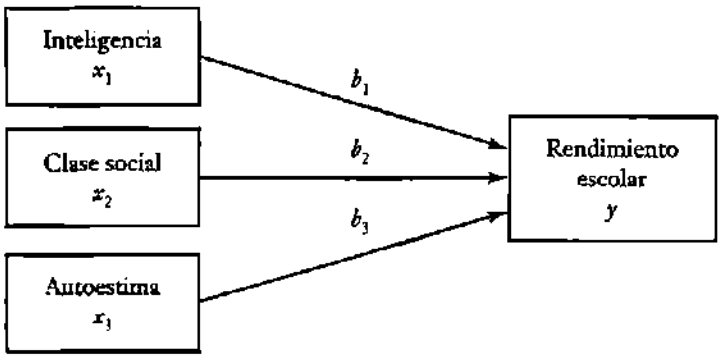
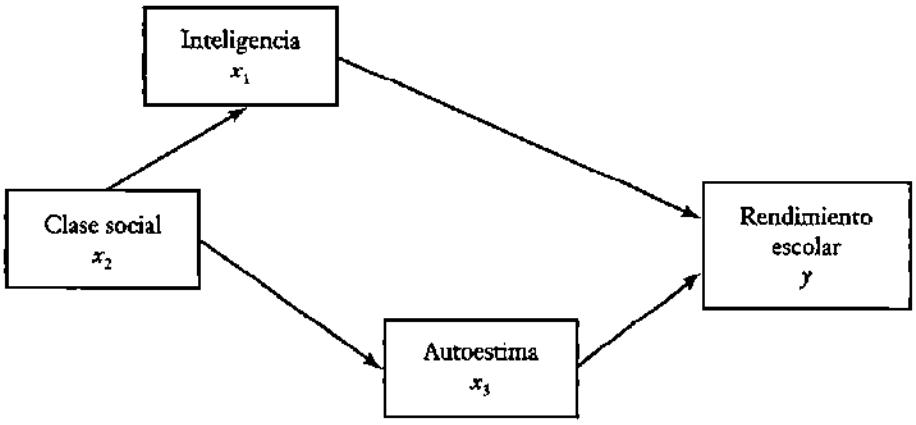


FIGURA 5.10



variable dependiente, se especifica en efecto, que las tres variables independientes afectan en forma directa a la variable dependiente. Esto se llama un problema de regresión múltiple directa (ver abajo) en el que k(=3) variables independientes influyen mutuamente en una variable dependiente. Este enfoque también ha cambiado en forma dramática en la década pasada. Los investigadores están ahora preparados para hablar de y probar tanto influencias directas como indirectas. Un modelo alternativo y un diagrama de ruta analítico se presentan en la figura 5.10. Aquí la Inteligencia y la Autoestima influyen en el Rendimiento Escolar en forma directa, pero la clase Social, no. En su lugar, influye de manera indirecta en el Rendimiento Escolar *a través* de la Inteligencia y la Autoestima, lo cual es un concepto bastante diferente.

Relaciones múltiples y regresión

La situación de investigación descrita en la figura 5.9 es un problema de regresión múltiple: k (=3) variables independientes influyen de forma mutua y simultánea en una variable dependiente. Más adelante se mostrará cómo se resuelve un problema de esta naturaleza. (El método es técnicamente complejo aunque conceptualmente simple, pero nos costará algo de trabajo.) Por ahora, el problema es encontrar inicialmente la relación entre las tres variables independientes, tomadas de forma simultánea, y la variable dependiente. El segundo punto consiste en determinar en qué medida cada variable independiente, x_1, x_2 y x_3, influye en la variable dependiente, y. Aunque es ahora mucho más complejo, el problema aún es una relación, un conjunto de pares ordenados.

Lo que el método hace en esencia —y bellamente— es encontrar la mejor combinación posible de x_1, x_2 y x_3 dada y, así como las relaciones entre las cuatro variables, de tal forma que la correlación entre la combinación de las tres variables y y sea máxima. En el problema mostrado en la figura 5.9, la regresión múltiple encuentra los valores de b_1, b_2 y b_3 tales que hagan la correlación entre x_1, x_2 y x_3 tomadas juntas, y y tan alta como sea posible. (El estudiante de matemáticas reconocerá que se trata de un problema de cálculo.) Los pesos de b, llamados pesos de regresión o coeficientes, se usan entonces con las tres variables para predecir la variable dependiente, y. El método, en efecto, crea una nueva variable que es una combinación de x_1, x_2 y x_3. Llamemos a esta variable y'. Así, la

correlación múltiple es entre y, la variable dependiente observada, y y', la variable dependiente predicha a partir del conocimiento de x_1, x_2 y x_3.

El lector que está pendiente y atento observará que las relaciones y correlaciones son simétricas: con frecuencia no importa cuál es la variable dependiente y cuál la independiente. En el análisis de regresión, sin embargo, sí implica una diferencia en tanto que la regresión es asimétrica. Nosotros decimos: si x entonces y o: si x_1, x_2 y x_3, entonces y. Muchos autores hablan de "análisis causal", en especial cuando se refieren a problemas como el de las figuras 5.9 y 5.10. Nosotros preferimos evitar las palabras *causa y causal* porque son ideas excesivamente difíciles —por ejemplo, ¿qué es una causa?— y porque no resulta necesario usarlas. Comrey y Lee (1992, p. 338) establecen que "las inferencias causales no pueden hacerse con certeza. Lo más que se puede decir es que los datos son consistentes con la inferencia causal propuesta..." Así, podemos operar de manera adecuada con enunciados condicionales, aunque no siempre con facilidad.[3]

La regresión en otras palabras, trata con relaciones, aunque en general es un camino de una sola vía, de variables independientes a dependientes. Para anticipar una discusión posterior, veamos una ecuación de regresión:

$$Y = a + b_1X_1 + b_2X_2$$

Si ignoramos a —que no es importante para el argumento— podemos ver que Y es la suma de X_1 y X_2, cada una ponderada por su propia b. Cuando resolvemos la ecuación para las bs (y por supuesto, la a), las usamos para producir un resultado de Y' para cada persona en la muestra. Y y Y' (conserve en la mente que Y y Y' representan valores para cada persona en la muestra) son entonces un conjunto de pares ordenados y, por lo tanto, una relación. La correlación entre ellas es meramente un coeficiente de correlación ordinario, r. Pero se le denomina R y se le llama coeficiente múltiple de correlación o coeficiente de correlación múltiple. Más adelante se examinará el uso e interpretación de la regresión múltiple, el coeficiente de correlación múltiple, así como los pesos de la regresión con mayor detalle y con ejemplos de investigación reales. En este momento el asombro natural del lector a partir de los misterios supuestos del pensamiento multivariado deben haberse disipado y reemplazado por admiración y quizá un poco de sobrecogimiento y excitación por lo atractivo y poderoso de estas ideas y métodos.

RESUMEN DEL CAPÍTULO

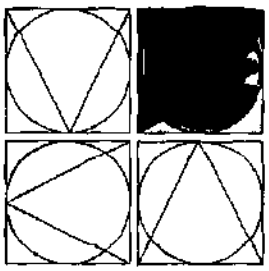
1. Las relaciones son la esencia del conocimiento. Casi toda ciencia busca y estudia relaciones.
2. Las relaciones en la ciencia se dan entre clases o conjuntos de objetos.
3. Las relaciones pueden expresarse como conjuntos de pares ordenados.
4. Los pares ordenados son conjuntos de elementos con un orden fijo de aparición.
5. Hay dos conjuntos especiales: dominio e imagen de una relación.
6. La función es un tipo especial de relación que conecta elementos del dominio y de la imagen.
7. Los miembros de un conjunto se mapean en los miembros del otro conjunto por medio de una regla de correspondencia.

[3] El lenguaje está saturado con palabras que implican causa, por ejemplo, "influencia" y "dependencia". Evitamos al máximo el uso de expresiones causales por la sola razón de que nunca es posible decir sin ambigüedades que una cosa causa otra. De forma más pragmática, no necesitamos la palabra o concepto "causa"; los enunciados condicionales de si p, entonces q son suficientes para los propósitos científicos.

8. La regla de correspondencia es una receta o fórmula que muestra cómo se mapea los objetos.
9. Formas de estudiar las relaciones:
   a) Gráficos (de dos dimensiones)
   b) Tablas
   c) Gráficos y correlación (donde la correlación es un valor estadístico/numérico).
10. La regresión múltiple es un método estadístico que relaciona una variable dependiente a una combinación lineal de una o más variables independientes. Este procedimiento incluso puede decirle al investigador en qué medida cada variable independiente explica o se relaciona a la variable dependiente.

SUGERENCIAS DE ESTUDIO

1. Los análisis sobre relaciones parecen estar confinados a los textos de matemáticas. El mejor análisis que hemos encontrado, aunque abstracto y algo difícil, está en Kershner y Wilcox (1974).
2. A continuación se presentan seis ejemplos de relaciones. Suponga que el primer conjunto es el dominio y el segundo, la imagen. ¿Por qué son todas estas relaciones?
   a) Páginas del libro y número de las páginas
   b) Números de los capítulos y páginas de un libro
   c) Encabezados o categorías en una tabla de población y las cantidades de población en un informe de censos
   d) Niños de una clase de 3er. grado y sus puntuaciones en una prueba estandarizada de rendimiento.
   e) $Y = 2x$
   f) $Y = a + b_1X_1 + b_2X_2$
3. Un investigador educativo ha estudiado la relación entre ansiedad y rendimiento escolar. Exprese la relación en el lenguaje de conjuntos.
4. Suponga que desea estudiar las relaciones entre las siguientes variables: inteligencia, estatus socioeconómico, necesidad de logro y rendimiento escolar. Elabore dos modelos alternativos que "expliquen" el rendimiento escolar. Dibuje diagramas de ruta para cada uno de los modelos.
5. Determine cuáles de las siguientes relaciones son funciones:
   a) $(\neq Q, R \neq, \neq S, T \neq)$
   b) $(\neq w, j \neq, \neq k, l \neq, \neq p, q \neq)$
   c) $(\neq a, b \neq, \neq 102, 103 \neq, \neq a, c \neq)$
   d) Dado $A = \{a, b, c\}$ y $B = \{4, 5, T\}$, ¿el producto cartesiano cruzado $A \times B$ es una función? Explique por qué sí o por qué no.



CAPÍTULO 6

VARIANZA Y COVARIANZA

- **CÁLCULO DE MEDIAS Y VARIANZAS**
- **TIPOS DE VARIANZA**
  - **Varianza poblacional y muestral**
  - **Varianza sistemática**
  - **Varianza entre grupos (experimental)**
  - **Varianza del error**
    - *Un ejemplo de la varianza sistemática y varianza del error*
    - *Una demostración sustractiva: remoción de la varianza entre grupos de la varianza total*
    - *Una recapitulación de la remoción de la varianza entre grupos de la varianza total*
- **COMPONENTES DE LA VARIANZA**
- **COVARIANZA**
  - **Anexo computacional**

Para estudiar problemas científicos y responder preguntas científicas debemos estudiar las diferencias entre fenómenos. En el capítulo 5 examinamos relaciones entre variables; en un sentido, estudiábamos similitudes. Ahora nos concentraremos en las diferencias, ya que sin diferencias ni variaciones no hay manera técnica de determinar las relaciones entre variables. Si queremos estudiar la relación entre raza y aprovechamiento, por ejemplo, estamos indefensos si tan sólo contamos con medidas de aprovechamiento de los niños caucásicos estadounidenses. Debemos tener medidas de aprovechamiento de más de una raza. En síntesis, es necesario que la raza varíe; debe tener varianza. Se requiere explorar el concepto de varianza de manera analítica y con cierta profundidad. Para hacerlo adecuadamente es necesario también retirar algo de la crema de la leche de las estadísticas.

Estudiar conjuntos de números como tales resulta pesado. En general es necesario reducir los conjuntos en dos formas: 1) por medio del cálculo de promedios o medidas de tendencia central, y 2) por medio del cálculo de medidas de variabilidad. La medida de tendencia central usada en este libro es la media. La medida de variabilidad más usada es la *varianza*. Ambas clases de medidas sintetizan conjuntos de puntuaciones, pero de diferentes maneras. Ambas constituyen "resúmenes" de conjuntos completos de puntuaciones y expresan dos importantes facetas de dichos conjuntos: 1) su tendencia o promedio central,

y 2) su variabilidad. Resolver problemas de investigación sin estas medidas es en extremo difícil. Empezaremos el estudio de la varianza con algunos cálculos simples.

Cálculo de medias y varianzas

Tomemos el conjunto de números $X = \{1, 2, 3, 4, 5\}$. La media se define:

$$M = \frac{\Sigma X}{n} \tag{6.1}$$

n equivale al número de casos en el conjunto de puntuaciones; Σ significa "la suma de" o "súmelos" y X representa a cualquiera de las puntuaciones (cada puntuación es una X). Entonces, la fórmula se lee: "sume las puntuaciones y divida entre el número de casos en el conjunto".

$$M = \frac{1 + 2 + 3 + 4 + 5}{5} = \frac{15}{5} = 3$$

La media del conjunto X es 3. En este libro "M" representará la media. Otros símbolos que se usan con frecuencia son $\bar{X}$ y μ.

El cálculo de la varianza, aunque no tan sencilla como el de la media, aún es simple. La fórmula es:

$$V = \frac{\Sigma x^2}{n} \tag{6.2}$$

donde V es la varianza; n y Σ son lo mismo que en la ecuación 6.1. Σx^2 se denomina suma de cuadrados (esto necesita algo de explicación). Las puntuaciones se enlistan en una columna:

	X	x	x^2
	1	−2	4
	2	−1	1
	3	0	0
	4	1	1
	5	2	4
ΣX:	15		
M:	3		
Σx^2:			10

En este cálculo, x es una desviación de la media. Se define como:

$$x = X - M \tag{6.3}$$

Así, para obtener x tan sólo reste de X la media de todas las puntuaciones. Por ejemplo si $X = 1$, $x = 1 - 3 = -2$; en el caso de $X = 4$, $x = 4 - 3 = 1$; etcétera. Esto se hizo en la tabla anterior. La ecuación 6.2, sin embargo, indica que se elevó al cuadrado cada x. Esto también se hizo antes (recuerde que el cuadrado de un número negativo siempre es positivo). En otras palabras, $\sum x^2$ señala que hay que restar la media a cada puntuación para obtener x, elevar al cuadrado cada x para obtener x^2, y entonces sumar todas las x^2. Por último el promedio de las x^2 se genera al dividir $\sum x^2$ entre n, el número de casos. $\sum x^2$, la *suma de cuadrados*, es un estadístico muy importante que usaremos con frecuencia.

La varianza en el presente caso es

$$V = \frac{(-2)^2 + (-1)^2 + (0)^2 + (1)^2 + (2)^2}{5} = \frac{4 + 1 + 0 + 1 + 4}{5} = \frac{10}{5} = 2$$

"V" representará la varianza en este libro. Otros símbolos comúnmente usados son σ^2 y s^2. El primero también se llama valor poblacional; el segundo es un valor muestral. N se usa para representar el número total de casos en una muestra total o en una población. (Se definirá "muestra" y "población" en un capítulo posterior.) n se usa para una submuestra o subconjunto de U de una muestra total. Los subíndices apropiados se anexarán y explicarán conforme sea necesario. Por ejemplo, si queremos indicar el número de elementos en el conjunto A, subconjunto de U, podemos escribir n_A o n_a. De manera similar agregaremos subíndices para x, V, etcétera. Cuando se use doble subíndice, como r_{xy}, el significado por lo general resultará obvio.

La varianza se llama también *cuadrado medio* (cuando se calcula de una forma ligeramente diferente). Se llama así porque es evidente que equivale a la media de las x^2. Es claro que no es difícil calcular la media y la varianza.[1]

La pregunta es: ¿Por qué calcular la media y la varianza? El fundamento para el cálculo de la media se explica fácilmente. La media expresa el nivel general, el centro de gravedad, de un conjunto de medidas. Es un buen representante del nivel de las características o rendimiento de un grupo. También posee ciertas propiedades estadísticas deseables y es el estadístico más frecuente en las ciencias del comportamiento. En gran parte de este tipo de investigación, por ejemplo, se compara la media de diferentes grupos experimentales para estudiar relaciones como se señaló en el capítulo 5. Podemos probar la relación entre climas organizacionales y productividad, por ejemplo. Pudimos usar tres tipos de climas y estar interesados en conocer qué clima tiene el mayor efecto en la productividad. En tales casos, las medias por lo general se comparan. Por ejemplo, de tres grupos, cada uno operando bajo alguno de los tres climas A_1, A_2 y A_3, ¿cuál tiene la mayor media en, digamos, una medida de productividad?

La base lógica para los cálculos y uso de la varianza en investigación es más difícil de explicar. En el caso usual de puntuaciones ordinarias, la varianza es una medida de dispersión del conjunto de puntuaciones: nos dice qué tanto se dispersan los valores. Si un grupo de alumnos es muy heterogéneo en cuanto a rendimiento en lectura, la varianza de sus puntuaciones en este rubro será muy grande comparada con la varianza de un grupo que es homogéneo en este aspecto. La varianza, entonces, es una medida de la dispersión de las puntuaciones; describe la medida en que las puntuaciones difieren entre sí. Para propósi-

[1] El método para calcular la varianza usado en este capítulo difiere de otros que se usan de ordinario. De hecho, el presentado aquí es impracticable en la mayoría de las situaciones. Nuestro propósito no es aprender estadística como tal. Lo importante es ir tras las ideas básicas. Se han construido métodos para realizar cálculos, ejemplos y demostraciones para ayudar en esta búsqueda de ideas básicas.

tos descriptivos, por lo general se usa la raíz cuadrada de la varianza, y se denomina *desviación estándar*. Algunas propiedades matemáticas, sin embargo, hacen a la varianza más útil en investigación. Se sugiere que el estudiante complemente este tópico con las secciones apropiadas de un texto elemental de estadística (véase Comrey y Lee, 1995). No es posible discutir en este libro todas las facetas del significado e interpretación de medias, varianzas y desviaciones estándar. El resto de este capítulo y las partes posteriores de este libro explorarán otros aspectos del uso del estadístico varianza.

Tipos de varianza

La varianza viene en varias formas. Cuando lea literatura especializada y de investigación, con frecuencia se topará con el término, algunas veces con algún adjetivo calificativo, otras no. Para entender la literatura es necesario tener una idea clara de las características y propósitos de las diferentes varianzas. Para diseñar y hacer investigación, uno debe tener un profundo entendimiento del concepto de varianza así como un dominio amplio de los conceptos y manipulaciones estadísticas de la varianza.

Varianza poblacional y muestral

La *varianza poblacional* es la varianza de U, un universo o población de medidas. En general, se usan símbolos griegos para representar los parámetros o medidas poblacionales. Para la varianza poblacional, se usa el símbolo σ^2 (sigma cuadrada). El símbolo σ se utiliza para la desviación estándar poblacional. La media poblacional es μ (mu). Si se conocen todas las medidas de un conjunto universal definido, U, entonces se conoce la varianza. Con frecuencia, sin embargo, la totalidad de medidas de U no están disponibles. En tales casos, la varianza se estima al calcular la varianza de una o más de las muestras de U. Una buena parte de la energía estadística se dirige a esta importante actividad. Quizá surja una pregunta: ¿Qué tanto varía la inteligencia de los ciudadanos de Estados Unidos? Se trata de una pregunta poblacional o de U. Si hubiera una lista completa de todos los millones de gente en Estados Unidos —y también un listado completo de las puntuaciones de pruebas de inteligencia de estas personas— la varianza podría calcularse en forma simple, aunque pesada. Tal lista no existe. Entonces se prueban las muestras, muestras representativas, de estadounidenses y se calculan las medias y varianza. Se utilizan muestras para estimar la media y varianza de toda la población. Estos valores estimados se llaman estadísticos (en la población se denominan parámetros). La media muestral se denota por el símbolo M y la varianza muestral por SD^2 o s^2. Muchos libros de estadística usan $\bar{X}$ (X-barra) para representar la media muestral.

La varianza de muestreo es la varianza de los estadísticos calculados a partir de muestras. Las medias de cuatro muestras aleatorias tomadas de una población diferirán. Si el muestreo es aleatorio y las muestras lo suficientemente grandes, las medias no deben variar mucho; esto es, la *varianza de las medias* deberá ser relativamente pequeña.[2]

[2] Por desgracia, en la mayor parte de la investigación actual sólo está disponible una muestra —y con frecuencia es pequeña—. Podemos, sin embargo, estimar la varianza muestral de las medias por medio de *la varianza estándar de la media*. (El término "error estándar de la media" se usa con frecuencia y es la raíz cuadrada de la varianza estándar de la media.) La fórmula es $V_M = V_s/n_s$ donde V_M es la varianza estándar de la media, V_s es la varianza de la muestra, y n_s es el tamaño de la muestra. Observe una conclusión importante que puede rescatarse

Varianza sistemática

Quizás la forma más general de clasificar la varianza es en varianza sistemática y *varianza del error*. La *varianza sistemática* es la variación en las medidas debidas a influencias conocidas o desconocidas que "causan" que las puntuaciones se inclinen a una dirección más que a otra. Cualquier influencia natural o generada por el hombre que cause que los eventos sucedan de una forma predecible son influencias sistemáticas. Las puntuaciones de rendimiento en pruebas de niños de una adinerada escuela suburbana tienden a ser sistemáticamente más altas que las de los alumnos de una escuela en un barrio pobre. Los maestros expertos pueden de manera sistemática influir en el aprovechamiento de los niños, en contraste con aquellos que reciben una enseñanza deficiente.

Hay muchas causas de la varianza sistemática. Los científicos buscan separar aquellas que les interesan de aquellas que no. También intentan separar la varianza aleatoria de la varianza sistemática. De hecho, la investigación puede definirse de forma precisa y técnica, como el estudio controlado de varianzas.

Varianza entre grupos (experimental)

Un tipo importante de varianza sistemática en la investigación es la varianza entre grupos o varianza experimental. La *varianza entre grupos* o *experimental*, como su nombre lo indica, es aquella que refleja diferencias sistemáticas entre *grupos* de medidas. La varianza discutida previamente como varianza de puntuaciones refleja las diferencias entre individuos en un grupo. Podemos decir por ejemplo, que, con base en la evidencia y pruebas actuales, la varianza en la inteligencia de una muestra aleatorizada de niños de 11 años de edad es de cerca de 225 puntos. (Se obtiene al elevar al cuadrado la desviación estándar derivada del manual de una prueba. La desviación estándar de la Prueba California de Madurez Mental para niños de 11 años de edad, por ejemplo, es de alrededor de 15, y $15^2 = 225$.) Esta cantidad es un estadístico que nos indica la medida en que los individuos difieren uno de otro.

La varianza entre grupos, por otro lado, es aquella que se debe a las diferencias entre *grupos* de individuos. Si se mide el aprovechamiento de niños de la región norte y de la región sur en escuelas comparables, habría diferencias entre los grupos del norte y del sur. Los grupos al igual que los individuos difieren o varían, y es posible y apropiado calcular la varianza entre estos grupos.

La varianza entre grupos y la experimental son en esencia iguales. Ambas provienen de las diferencias entre grupos. La varianza entre grupos es un término que abarca todos los casos de diferencias sistemáticas entre grupos, tanto experimentales como no experimentales. La varianza experimental con frecuencia se asocia con la varianza originada por la manipulación activa de las variables independientes por parte de los investigadores.

He aquí un ejemplo de varianza entre grupos —en este caso, varianza experimental—. Suponga que un investigador prueba la eficacia relativa de tres diferentes clases de reforzamiento en el aprendizaje. Después de reforzar a los tres grupos de sujetos de forma diferencial, el experimentador calcula la media de los grupos. Suponga que son 30, 23 y 19. La media de las tres medias es 24, y calculamos la varianza *entre las medias* o *entre los grupos*:

---

de esta ecuación: si se incrementa el tamaño de la muestra, V_M disminuye. En otras palabras, para hacer que la muestra sea más confiable y cercana a la media poblacional, haga la n más grande. Por el contrario, mientras más pequeña sea la muestra, más riesgosa se hace la estimación (ver las sugerencias de estudio 5 y 6).

	X	x	x^2
	30	6	36
	23	−1	1
	19	−5	25
ΣX:	72		
M:	24		
Σx^2:			62

$$V_r = \frac{62}{3} = 20.67$$

En este experimento, se presume que los diferentes métodos de reforzamiento tienden a "sesgar" las puntuaciones en un sentido u otro, lo cual es, por supuesto, el propósito del investigador. El objetivo del método A es aumentar todas las puntuaciones, relativas al aprendizaje de un grupo experimental. El experimentador puede creer que el método B no tendrá efecto en el aprendizaje y que el método C tendrá un efecto depresor. Si el experimentador está en lo correcto, las puntuaciones del método A tenderán a aumentar, mientras que las del método C tenderán a disminuir. Así, las puntuaciones de los grupos, como un todo —y por supuesto sus medias— difieren de forma sistemática. El reforzamiento es una variable *activa*, manipulada a propósito por el experimentador con la intención consciente de "sesgar" las puntuaciones en una forma diferencial. Prokasy (1987), por ejemplo, ayudó a consolidar este punto al resumir el número de variaciones de reforzamiento en el paradigma pavloviano en el estudio de respuestas de tipo esquelético. Así, cualesquier variables manipuladas por el experimentador están íntimamente asociadas con la varianza sistemática. Camel, Withers y Greenough (1986) dieron a su grupo experimental de ratas diferentes grados de experiencia temprana —ambiental (experiencias enriquecidas tal como una jaula grande con otras ratas y oportunidades para explorar), y al grupo control una condición de experiencia reducida (aislamiento y jaulas individuales)—. Ellos intentaron de forma deliberada construir una varianza sistemática en sus resultados medidos (patrón y número de ramificaciones dendríticas [las dendritas son las estructuras ramificadas de una neurona]). La idea básica atrás del famoso "diseño clásico" de investigación científica en el que se usan grupos experimental y control es que a través del control y manipulación cuidadosos, se hacen variar de forma sistemática las medidas del resultado del grupo experimental (también llamadas "medidas de criterio") para aumentar o disminuir, mientras que las medidas del grupo control se mantienen por lo general al mismo nivel. La varianza por supuesto, es entre los dos grupos, es decir, se hace que ambos grupos difieran. Por ejemplo, Braud y Braud (1972) manipularon grupos experimentales en una forma por demás inusual. Entrenaron ratas de un grupo experimental para elegir el mayor de dos círculos en una tarea de elección; el grupo control de ratas no recibió entrenamiento. Después se inyectó en los cerebros de dos nuevos grupos de ratas extractos de cerebro de los animales de ambos grupos. Desde el punto de vista estadístico intentaron aumentar la varianza entre grupos y tuvieron éxito: ¡Los sujetos de los nuevos "grupos experimentales", superaron a los del nuevo "grupo control" en la tarea de elegir el mayor círculo en la misma tarea de elección!

Esto es claro y fácil de ver en experimentos. En la investigación no experimental, cuando existen diferencias entre los grupos estudiados, no siempre es tan sencillo y directo el visualizar la varianza entre grupos que uno estudia. Pero la idea es la misma. El principio puede establecerse de forma algo diferente: a mayor diferencia entre grupos, más puede asumirse que la o las variables independientes han operado. Si hay una pequeña diferencia entre grupos, debe presumirse que la o las variables independientes *no* han operado. En otras palabras, o bien sus efectos son demasiado débiles para ser aparentes, o

bien diversas influencias se han cancelado mutuamente. Así, juzgamos los efectos de las variables independientes manipuladas o que han trabajado en el pasado a través de la varianza entre grupos. Ya sea que se hayan manipulado o no las variables independientes, el principio es el mismo.

Para ilustrar el principio, usaremos el bien estudiado problema del efecto de la ansiedad en el aprovechamiento escolar. Es posible manipular la ansiedad al contar con dos grupos experimentales e inducir ansiedad en uno y en el otro no. Esto se puede lograr al aplicar a cada grupo la misma prueba con diferentes instrucciones. Decimos a los miembros de un grupo que su calificación dependerá por completo de la prueba, mientras que al otro grupo le decimos que el examen no tiene una importancia en particular y que el resultado no afectará sus notas. Por otro lado, la relación entre ansiedad y aprovechamiento puede también estudiarse al comparar grupos de individuos en los cuales se presupone que diferentes circunstancias ambientales y psicológicas han actuado para producir ansiedad. (Por supuesto, se asume que la ansiedad inducida de manera experimental y la ansiedad preexistente —la variable de estímulo y la variable organísmica— no son iguales.) Guida y Ludlow (1989) condujeron un estudio para probar la hipótesis de que diferentes circunstancias ambientales y psicológicas actúan para producir diferentes niveles de ansiedad al resolver exámenes. Estos investigadores hipotetizaron que los estudiantes en la cultura de Estados Unidos mostrarían un nivel más bajo de ansiedad ante las pruebas que los alumnos de la cultura chilena. Para usar el lenguaje de este capítulo, los investigadores hipotetizaron una mayor varianza entre grupos que la esperada debido al azar a causa de la diferencia entre las condiciones ambientales, educativas y psicológicas chilenas y estadounidenses. (Se encontró apoyo para la hipótesis. Los estudiantes chilenos exhibieron un mayor nivel de ansiedad al resolver exámenes que los alumnos de Estados Unidos. Sin embargo, al considerar sólo los grupos socioeconómicos más bajos de ambas culturas, los aprendices de Estados Unidos tuvieron un mayor nivel de ansiedad que los chilenos.)

Varianza del error

La *varianza del error* es la fluctuación o variación de medidas que no se pueden explicar. Las fluctuaciones de las mediciones en la variable dependiente en un estudio de investigación donde todos los participantes fueron tratados de igual forma se considera varianza del error. Algunas de estas variaciones se deben al azar: en este caso, la varianza del error es varianza aleatoria. Consiste en la variación en las medidas debida a fluctuaciones usualmente pequeñas y autocompensadoras —ahora aquí, ahora allá; ahora arriba, ahora abajo—. La varianza muestral antes discutida, por ejemplo, es varianza del error o aleatoria.

Hagamos una breve digresión en tanto que en este capítulo y el siguiente se usa el concepto de "azar" o "aleatoriedad". Las ideas de aleatoriedad y aleatorización se discutirán con mucho más detalle en el capítulo 8. Por ahora, ***aleatoriedad*** significa que no hay una forma conocida de describir correctamente o explicar los eventos y sus resultados en términos de lenguaje. En otras palabras, los eventos azarosos no pueden predecirse. Una muestra aleatoria es un subconjunto de un universo; se selecciona a sus miembros de tal forma que cada miembro del universo tiene igual posibilidad de ser elegido. Ésta es otra forma de decir que si los miembros son seleccionados aleatoriamente, no hay forma de predecir qué miembro será elegido en cada oportunidad, al mantener constantes las demás condiciones.

Sin embargo, no debe pensarse que la varianza aleatoria es la única fuente posible de varianza del error. La varianza del error puede constar también de otros componentes como lo señaló Barber (1976). Todo lo que pudiera estar incluido en el término "varianza del error" puede incluir errores de medición en el instrumento usado, errores de procedi-

miento llevados a cabo por el investigador, registro erróneo de las respuestas y la expectativa que el investigador tiene de los resultados. Es posible que "sujetos iguales" difieran en la variable dependiente porque uno de ellos puede estar experimentando un funcionamiento fisiológico o psicológico distinto al momento en que las mediciones fueron tomadas.

Para regresar a nuestra discusión principal, puede decirse que la varianza del error es la varianza en las mediciones debida a ignorancia. Imagine un gran diccionario en el que todo en el mundo —cada ocurrencia, evento, pequeña cosa, asunto importante— se presenta con todo detalle. Para entender cualquier evento que ha ocurrido, sucede ahora o pasará, todo lo que se necesita hacer es mirar el diccionario. Con él es evidente que no hay posibilidad de ocurrencias azarosas. Todas las cosas están explicadas. Es decir, no hay varianza del error; todo es varianza sistemática. Por desgracia (o más bien, por fortuna) no contamos con ese diccionario. Así, muchos eventos y ocurrencias no pueden explicarse. Una gran parte de la varianza elude la identificación y el control. Ello constituye la varianza del error mientras se nos escape su identificación y control.

Aunque parezca un poco extraño y aun bizarro, este modo de razonamiento es útil, mientras sepamos que una parte de la varianza del error del hoy puede no serlo mañana. Suponga que conducimos un experimento sobre la enseñanza de solución de problemas en el cual asignamos alumnos a tres grupos de forma aleatoria. Al terminar el experimento, estudiamos las diferencias entre los tres grupos para ver si la enseñanza tuvo algún efecto. Sabemos que las puntuaciones y medias de los grupos siempre mostrarán fluctuaciones menores, a veces un punto o dos o tres más, y en otras uno o dos o tres menos, que probablemente nunca se controlarán. Algo siempre hace variar así las puntuaciones. De acuerdo a la postura que se analiza, no fluctúan porque sí: es probable que no exista la "aleatoriedad absoluta". Desde una postura determinista, debe haber alguna causa (o causas) para las fluctuaciones. Así es, podemos identificar algunas y quizá controlarlas. Cuando hacemos esto, sin embargo, tenemos varianza sistemática.

Descubrimos, por ejemplo, que el género "causa" la fluctuación de la puntuaciones, en tanto que hay hombres y mujeres mezclados en los grupos experimentales. (Por supuesto, hablamos de forma figurativa. Es evidente que el género no hace que las puntuaciones fluctúen.) Así que conducimos el experimento y controlamos el género al incluir, por ejemplo, sólo a hombres. Las puntuaciones aún varían, pero menos. Entonces retiramos otra supuesta causa de las alteraciones: la inteligencia. La puntuación todavía fluctúa, aunque algo menos. Continuamos retirando otras fuentes de varianza. Así, controlamos la varianza sistemática al tiempo de identificar y controlar cada vez más la varianza desconocida de forma gradual.

Ahora observe que antes de controlar o retirar estas varianzas sistemáticas, antes de "saber" acerca de ellas, tendríamos que haberlas denominado "varianza del error" —en parte por ignorancia y en parte por nuestra incapacidad para controlar o para hacer algo acerca de ella—. Podemos hacer esto una y otra vez, y aún así habrá varianza residual: finalmente nos rendimos, ya no "sabemos" nada más; hemos hecho todo lo posible y todavía queda varianza. Así, una definición práctica de varianza del error sería: La *varianza del error* es aquella que persiste en un conjunto de medidas después de que todas las fuentes conocidas de varianza sistemática se han retirado de dichas medidas. Esto es tan importante que merece un ejemplo numérico.

Un ejemplo de varianza sistemática y varianza del error

Supongamos que estamos interesados en conocer si la cortesía al frasear instrucciones para una tarea afecta la memoria de las palabras amables. Llamamos "cortesía" y "descor-

tesía" a la variable A dividida en A_1 y A_2 (esta idea es de Holtgraves, 1997). Se asigna a los estudiantes al azar a dos grupos. Se define al azar qué grupo recibe el tratamiento A_1 y cuál A_2. En este experimento, los alumnos en A_1 recibieron instrucciones fraseadas sin cortesía, tales como, "usted debe escribir el nombre completo de cada estado que recuerde". Los estudiantes en A_2, por su parte, leyeron instrucciones con igual significado pero presentadas en forma cortés: "sería útil que usted escribiera el nombre completo de cada estado que recuerde". Después de leer las instrucciones, los sujetos tuvieron una tarea distractora consistente en recordar los 50 estados de la Unión Americana. Después se les aplicó una prueba de memoria de reconocimiento. Se usa este examen para determinar el recuerdo general de todas las palabras corteses. Las puntuaciones fueron:

	A_1	A_2
	3	6
	5	5
	1	7
	4	8
	2	4
M	3	6

Las medias son diferentes; éstas varían. Hay varianza entre grupos. Al tomar la diferencia entre las medias con el valor aparente —más adelante lo veremos con más precisión— podemos concluir que la vaguedad en la forma de expresarse tuvo un efecto. Al calcular la varianza entre grupos como lo hicimos antes, tenemos:

		x	x^2
	3	1.5	2.25
	6	1.5	2.25
M:	4.5		
Σx^2:			4.50

$$V_e = \frac{4.5}{2} = 2.25$$

En otras palabras, calculamos la varianza entre grupos como antes lo hicimos para las cinco puntuaciones 1, 2, 3, 4 y 5. Tan sólo tratamos las dos medias como si fueran puntuaciones individuales, y seguimos adelante con un cálculo ordinario de varianza. La varianza entre grupos, V_e, es entonces, 2.25. Una prueba estadística apropiada mostraría que la diferencia entre las medias de ambos grupos es lo que se llama "estadísticamente significativa". (Su significado se analizará en otro capítulo.)[3] Resulta evidente que el uso de palabras corteses en las instrucciones ayudó a incrementar las puntuaciones de memoria de los estudiantes.

[3] El método de cálculo usado aquí no es el que se utiliza para una prueba de significancia estadística; aquí tiene fines pedagógicos. Observe también que la escasa cantidad de casos en los ejemplos y las cifras pequeñas tienen el objetivo de simplificar la demostración. Los datos reales de una investigación, por supuesto, son en general más complejos y requieren de muchos más casos. En el análisis de varianza real, la expresión correcta para la suma de cuadrados-entre es: $SS_e = n\Sigma x_e^2$. Por simplicidad pedagógica, sin embargo, hemos conservado solamente Σx_e^2, para reemplazarla después por SS_e.

Si acomodamos las diez puntuaciones en una columna y calculamos la varianza tendremos:

X	x	x^2
3	−1.5	2.25
5	.5	.25
1	−3.5	12.25
4	−.5	.25
2	−2.5	6.25
6	1.5	2.25
5	.5	.25
7	2.5	6.25
8	3.5	12.25
4	−.5	.25
M:	4.5	
Σx^2:		42.50

$$V_t = \frac{42.5}{10} = 4.25$$

Ésta es la varianza total, V_t. $V_t = 4.25$ contiene todas las fuentes de variación en las puntuaciones. Ya sabemos que una de ellas es la varianza entre grupos, $V_e = 2.25$. Ahora calculemos otra varianza. Lo logramos al calcular la varianza de A_1 sola y la varianza de A_2 sola, para después promediar ambas:

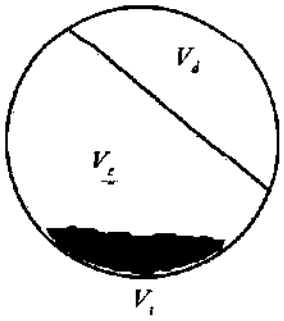
A_1	x	x^2	A_2	x	x^2
3	0	0	6	0	0
5	2	4	5	−1	1
1	−2	4	7	1	1
4	1	1	8	2	4
2	−1	1	4	−2	4
ΣX: 15			30		
M: 3			6		
Σx^2:		10			10

$$V_{A_1} = \frac{10}{5} = 2 \qquad V_{A_2} = \frac{10}{5} = 2$$

La varianza de A_1 es 2, y la varianza de A_2 es 2. El promedio es 2. Dado que cada una de las varianzas fue calculada por *separado* y después promediada, podemos llamar al promedio de la varianza calculada la "varianza dentro de grupos" (o intragrupos) y la denominamos V_d que significa varianza dentro de grupo, o varianza intragrupos. De ese modo, $V_d = 2$. *Esta varianza no se afecta por la diferencia entre las dos medias.* Se demuestra con facilidad al restar una constante de 3 a las puntuaciones de A_2; con ello, la media de A_2 es 3. Entonces, si se calcula la varianza de A_2, será la misma que antes: 2. Como es obvio la varianza intragrupos será la misma: 2.

Ahora escriba una ecuación: $V_t = V_e + V_d$. Esta ecuación indica que la varianza total está integrada por la varianza entre grupos y la varianza intragrupos. Pero, ¿así es? Sustituya

FIGURA 6.1



los valores numéricos: 4.25 = 2.25 + 2.00. Nuestro método funciona —y muestra, también, que las varianzas son aditivas (como se calculó)—.

Las ideas sobre varianza bajo discusión pueden quizás aclararse con un diagrama. En la figura 6.1 se divide un círculo en dos partes. Sea el área total del círculo la varianza total de las 10 puntuaciones o V_t. La porción más grande y sombreada representa la varianza entre grupos o V_e. El área más pequeña sin sombrear representa la varianza del error o varianza dentro de grupos o V_d. En el diagrama podemos ver que $V_t = V_e + V_d$. (Observe la similitud con el razonamiento de conjuntos y la operación de unión.)

V_t representa una medida de todas las fuentes de varianza V_e y a la medida de la varianza entre grupos (o una medida del efecto del tratamiento experimental). Pero ¿qué es V_d, la varianza intragrupos? Dado que, de la varianza total, hemos dado cuenta de una fuente conocida de varianza a través de la varianza entre grupos, podemos asumir que la varianza remanente se debe a factores derivados del azar que llamamos *varianza del error*. Pero, usted se preguntará seguramente si hay otras fuentes de varianza. ¿Qué hay de las diferencias individuales en inteligencia, género, etcétera? Ya que asignamos a los estudiantes a los grupos experimentales de manera azarosa, suponga que esas fuentes de varianza se distribuyen de igual forma, o casi igual, entre A_1 y A_2. Y, debido a la asignación al azar, no podemos aislar ni identificar otras fuentes de varianza. Esta varianza remanente se denomina *varianza del error*, con lo que sabemos muy bien que es probable que haya otras fuentes de varianza pero bajo el supuesto (y esperamos estar en lo correcto) de que están distribuidas de forma equitativa entre ambos grupos.

Una demostración sustractiva: remoción de la varianza entre grupos de la varianza total

Demostremos otra forma de retirar del conjunto original de puntuaciones la varianza entre grupos, a través de un sencillo procedimiento de sustracción. Primero, sea cada una de las medias de A_1 y A_2 igual a la media total; retiramos la varianza entre grupos. La media total es 4.5. (Véase arriba donde la media de las 10 puntuaciones fue calculada.) Segundo, ajuste cada puntuación individual de A_1 y A_2 al restar o sumar, según el caso, una constante apropiada. Dado que la media de A_1 es 3, sumamos 4.5 − 3 = 1.5 a cada una de las puntuaciones de A_1. La media de A_2 es 6 y 6 − 4.5 = 1.5 es la constante que será *restada* a cada una de las puntuaciones de A_2.

Estudie las puntuaciones "corregidas" y compárelas con las originales. Observe que variaron menos de lo que lo hicieron antes. Retiramos la varianza entre grupos, una por-

ción considerable de la varianza total. La varianza que permanece es la parte de la varianza total debida, presumiblemente, al azar. Calculamos la varianza de las puntuaciones "corregidas" de A_1, A_2 y la total, y observamos estos resultados sorprendentes:

Corrección: +1.5 −1.5

A_1	A_2
3 + 1.5 = 4.5	5 − 1.5 = 3.5
5 + 1.5 = 6.5	5 − 1.5 = 3.5
1 + 1.5 = 2.5	7 − 1.5 = 5.5
4 + 1.5 = 5.5	8 − 1.5 = 6.5
2 + 1.5 = 3.5	4 − 1.5 = 2.5
ΣX: 22.5	22.5
M: 4.5	4.5

A_1	x	x^2	A_2	x	x^2
4.5	0	0	4.5	0	0
6.5	2	4	3.5	−1	1
2.5	−2	4	5.5	1	1
5.5	1	1	6.5	2	4
3.5	−1	1	2.5	−2	4
ΣX: 22.5			22.5		
M: 4.5			4.5		
Σx^2:		10			10

$$V_{A_1} = \frac{10}{5} = 2 \qquad V_{A_2} = \frac{10}{5} = 2$$

La varianza intragrupos es la misma de antes. No se afecta por la operación de corrección. Como es evidente, la varianza entre grupos ahora es 0. Y, ¿qué sucede con la varianza total, V_t? Al calcularla, obtenemos $\Sigma x_t^2 = 20$, y $V_t = 20 + 10 = 2$. Así la varianza intragrupos es ahora igual a la varianza total. El lector debe estudiar este ejemplo con cuidado hasta que entienda con claridad lo que ha sucedido y *por qué*.

Aunque el ejemplo anterior quizá es suficiente para destacar los puntos esenciales, puede consolidarse la comprensión del estudiante sobre estas ideas básicas de la varianza si ampliamos el ejemplo al señalar otra fuente de varianza. El lector recordará que la varianza intragrupos contiene la variación debida a diferencias individuales. Ahora suponga que, en lugar de asignar a los alumnos a los dos grupos de forma aleatoria, los hemos apareado con base en su inteligencia, que está relacionada con la variable dependiente. Es decir, hemos asignado pares de miembros con aproximadamente igual puntuación en pruebas de inteligencia en ambos grupos. El resultado del experimento podría ser:

A_1	A_2
3	6
1	5
4	7
2	4
5	8
M: 3	6

Observe con detalle que la única diferencia entre este arreglo y el anterior es que el apareamiento ha causado que las puntuaciones covaríen. Las medidas de A_1 y A_2 ahora tienen casi igual orden. De hecho, el coeficiente de correlación entre ambos conjuntos de puntuaciones es de 0.90. Tenemos aquí otra fuente de varianza: la debida a diferencias individuales en inteligencia, que se refleja en el orden de los pares de medidas de criterio. (La relación precisa entre las ideas de orden y de apareamiento y sus efectos en la varianza se revisarán en otro capítulo. Por el momento el estudiante deberá tomar como acto de fe que aparear genera varianza sistemática.)

Esta varianza puede calcularse y extraerse como se hizo antes, excepto que hay una operación adicional. Primero iguale las medias de A_1 y A_2 y "corrija" las puntuaciones como antes. El resultado es:

Corrección:	+1.5	−1.5
	4.5	4.5
	2.5	3.5
	5.5	5.5
	3.5	2.5
	6.5	6.5
M:	4.5	4.5

Segundo, al igualar los renglones (haciendo la media de cada renglón igual a 4.5 y "corrigiendo" las puntuaciones de los renglones consecuentemente) encontramos los siguientes datos:

Corrección	A_1	A_2	Medias originales	Medias corregidas
0	4.5 + 0 = 4.5	4.5 + 0 = 4.5	4.5	4.5
+1.5	2.5 + 1.5 = 4.0	3.5 + 1.5 = 5.0	3.0	4.5
−1.0	5.5 − 1.0 = 4.5	5.5 − 1.0 = 4.5	5.5	4.5
+1.5	3.5 + 1.5 = 5.0	2.5 + 1.5 = 4.0	3.0	4.5
−2.0	6.5 − 2.0 = 4.5	6.5 − 2.0 = 4.5	6.5	4.5
M	4.5	4.5	4.5	4.5

Las medidas doblemente corregidas ahora muestran muy poca varianza. La varianza de las 10 puntuaciones doblemente corregidas es 0.10, en verdad muy pequeña. Por supuesto, no queda varianza entre grupos (columnas) o entre individuos (renglones) en las medidas. Después de una doble corrección, la varianza total completa es varianza del error. (Como se verá más adelante, cuando las varianzas tanto de columnas como de renglones se extraen de esta forma —aunque con un método más rápido y eficiente— se elimina la varianza intragrupos.)

Una recapitulación de la remoción de la varianza entre grupos de la varianza total

Ésta ha sido una larga operación. Una breve recapitulación de los principales puntos puede resultar útil. Cualquier conjunto de medidas tiene una varianza total. Si las medidas a

partir de las cuales se calcula esta varianza se han derivado de respuestas de seres humanos, siempre habrá al menos dos fuentes de varianza. Una se deberá a fuentes sistemáticas de variación como las diferencias individuales de los sujetos cuyas características o logros se han medido, y las diferencias entre los grupos o subgrupos involucrados en la investigación. La otra se derivará del error aleatorio, fluctuaciones de las medidas de las que no se puede dar cuenta en la actualidad. Las fuentes de varianza sistemática tienden a hacer que las puntuaciones se inclinen hacia una dirección u otra, lo que implica diferencias en las medias. Si el género es una fuente sistemática de varianza en un estudio sobre rendimiento escolar, por ejemplo, entonces la variable género tenderá a actuar de manera tal que las puntuaciones de aprovechamiento de las niñas tiendan a ser mayores que las de los niños. Las fuentes del error aleatorio, por otro lado, tienden a hacer que las medidas fluctúen en un sentido en cierto momento, y en otra forma al siguiente. Los errores aleatorios, en otras palabras, se autocompensan; tienden a balancearse (o cancelarse) uno al otro.

En cualquier experimento o estudio, la o las variables independientes constituyen una fuente de varianza sistemática —al menos así debiera ser—. El investigador "quiere" que los grupos experimentales difieran sistemáticamente y con frecuencia busca maximizar tal varianza al controlar o minimizar otras fuentes de varianza, tanto sistemáticas como del error. El ejemplo experimental de antes ilustra la idea adicional de que estas varianzas son aditivas, y debido a esta propiedad aditiva es posible analizar un conjunto de puntuaciones en sus varianzas sistemática y del error.

Componentes de la varianza

La discusión hasta este punto pudo convencer al estudiante de que cualquier varianza total tiene "componentes de varianza". El caso que acabamos de considerar, sin embargo, incluye un componente experimental debido a la diferencia entre A_1 y A_2, un componente resultado de diferencias individuales, y un tercer componente referido al error aleatorio. Ahora estudiaremos el caso de dos componentes de varianza experimental sistemática. Para hacerlo, sintetizamos las medidas experimentales, y las creamos a partir de componentes *conocidos* de varianza. En otras palabras, damos marcha atrás: empezamos con fuentes "conocidas" de varianza ya que no habrá varianza del error en las puntuaciones sintetizadas.

Tenemos una variable X con tres valores. Sea $X = \{0, 1, 2\}$. También tenemos otra variable, Y, con tres valores. Sea $Y = \{0, 2, 4\}$. X y Y son, entonces, fuentes *conocidas* de varianza. Asumimos una condición experimental idónea con dos variables independientes que actúan *concertadamente* para producir efectos en la variable dependiente, Z. Esto es, cada puntuación de X opera con cada puntuación de Y para producir una puntuación Z de la variable dependiente. Por ejemplo, la puntuación X, 0, no tiene influencia. La puntuación X, 1, opera con Y como sigue: $\{(1 + 0), (1 + 2), (1 + 4)\}$. De forma similar, la puntuación X, 2, opera con Y: $\{(2 + 0), (2 + 2)$ y $(2 + 4)\}$. Todo esto es fácil de visualizar si generamos Z de una forma clara.

El conjunto de las puntuaciones en una matriz de 3×3 (una matriz es cualquier conjunto o tabla rectangular de números) es el conjunto de las puntuaciones Z. El propósito de este ejemplo se perderá a menos que el lector recuerde que en la práctica no conocemos las puntuaciones de X y Y; sólo conocemos las puntuaciones de Z. En una situación experimental real, manipulamos o establecemos X y Y y esperamos que sean efectivas. Esto puede no resultar así. En otras palabras, los conjuntos $X = \{0, 1, 2\}$ y $Y = \{0, 2, 4\}$ nunca podrán conocerse así. Lo más que podemos hacer es estimar su influencia a partir de estimar la cantidad de varianza en Z debida a X y a Y.

		Y		
		0	2	4
X	0	0 + 0	0 + 2	0 + 4
	1	1 + 0	1 + 2	1 + 4
	2	2 + 0	2 + 2	2 + 4

=

	Z		
	0	2	4
0	0	2	4
1	1	3	5
2	2	4	6

Los conjuntos X y Y tienen las siguientes varianzas:

X	x	x^2
0	−1	1
1	0	0
2	1	1
ΣX:	3	
M:	1	
Σx^2:		2

Y	y	y^2
0	−2	4
2	0	0
4	2	4
ΣY:	6	
M:	2	
Σy^2:		8

$$V_x = \frac{2}{3} = .67 \qquad V_y = \frac{8}{3} = 2.67$$

El conjunto Z tiene la varianza como sigue:

Z	z	z^2
0	−3	9
2	−1	1
4	1	1
1	−2	4
3	0	0
5	2	4
2	−1	1
4	1	1
6	3	9
ΣZ:	27	
M:	3	
Σz^2:		30

$$V_z = \frac{30}{9} = 3.33$$

Ahora .67 + 2.67 = 3.34, o $V_z = V_x + V_y$, con los errores propios del redondeo.

Este ejemplo ilustra que, bajo ciertas condiciones, las varianzas operan de forma aditiva para producir las medidas experimentales que analizamos. Aunque el ejemplo es "puro" y por lo tanto irreal, es razonable. Es posible concebir a X y Y como variables independientes; pudieran ser el nivel de aspiración y las actitudes de los alumnos. Z puede ser el aprovechamiento verbal, una variable dependiente. El hecho de que las puntuaciones reales no se comportan exactamente de esta forma no modifica la idea. Se comportan así de manera aproximada. Planeamos la investigación para hacer este principio tan verídico como sea posible, y analizamos los datos como si fuera verdadero. ¡Y funciona!

Covarianza

La covarianza en realidad no es nada nuevo. Recordemos que en una discusión previa de conjuntos y correlación hablamos acerca de la relación entre dos o más variables análogas a la intersección de conjuntos. Sea X {0, 1, 2, 3}, un conjunto de medidas de actitud de cuatro niños. Sea $Y = \{1, 2, 3, 4\}$, un conjunto de medidas de aprovechamiento de los mismos niños, pero no en el mismo orden. Sea R un conjunto de pares ordenados de los elementos de X y Y, donde la regla de apareamiento sería: se parea cada medida de actitud con cada medida de aprovechamiento del sujeto, con la medida de actitud colocada primero. Suponga que resulta que $R = \{(0, 2), (1, 1), (2, 3), (3, 4)\}$. De acuerdo a nuestra definición previa de relación, este conjunto de pares ordenados constituye una relación, en este caso, entre X y Y. El resultado del cálculo de la varianza de X y de la varianza de Y es:

	X	x	x^2	Y	y	y^2
	0	−1.5	2.25	2	−.5	.25
	1	−.5	.25	1	−1.5	2.25
	2	.5	.25	3	.5	.25
	3	1.5	2.25	4	1.5	2.25
ΣX:	6			10		
M:	1.5			2.5		
Σx^2:			5			5

$$V_x = \frac{5}{4} = 1.25 \qquad V_y = \frac{5}{4} = 1.25$$

Ahora nos planteamos un problema. (Observe con atención en lo que sigue y que trabajaremos con desviaciones de la media, x y y, y no con las puntuaciones brutas originales.) Hemos calculado las varianzas de X y Y usando las x y y; esto es, las desviaciones de las medias respectivas de X y Y. Si podemos calcular la varianza de cualquier conjunto de puntuaciones, ¿no es posible calcular la relación *entre* dos conjuntos cualesquiera de puntuaciones en una forma similar? ¿Es concebible que podamos calcular la varianza de dos conjuntos de forma simultánea? Y si lo hacemos, ¿será ésta una medida de la varianza de ambos conjuntos unidos? ¿Será esta varianza también una medida de la relación entre los dos conjuntos?

Lo que deseamos es usar alguna operación estadística análoga a la operación intersección de conjuntos, $X \cap Y$. Para calcular la varianza de X o de Y, elevamos al cuadrado las desviaciones de la media, las x o y, y después las sumamos y promediamos. Una respuesta natural a nuestro problema es realizar una operación análoga en las x y y *juntas*. Para calcular la varianza de X, lo primero que hicimos fue: $(x_1 \cdot x_1), \ldots, (x_4 \cdot x_4) = x_1^2, \ldots, x_4^2$. ¿Por qué no hacemos esto tanto con las x como con las y, multiplicando los pares ordenados así: $(x_1 \cdot y_1), \ldots, (x_4 \cdot y_4)$? Así, en lugar de escribir Σx^2 o Σy^2 escribiríamos Σxy, como sigue:

x	y	=	xy
−1.5	−.5	=	.75
−.5	−1.5	=	.75
.5	.5	=	.25
1.5	1.5	=	2.25

$$\Sigma xy = 4.00$$

$$V_{xy} = CoV_{xy} = \frac{4}{4} = 1.00$$

Vamos a darle nombre a Σxy y a V_{xy}. Σxy se llama *producto cruzado* o la suma de los productos cruzados. V_{xy} es llamada *covarianza*, que denotaremos con CoV con los subíndices apropiados. Si calculamos la varianza de estos productos, simbolizados como V_{xy} o CoV_{xy}, obtendríamos 1.00, como se indicó antes. Este 1.00, entonces, puede considerarse como un índice de la relación entre ambos conjuntos. Pero se trata de un índice no satisfactorio porque su tamaño fluctúa con los márgenes y escalas de diferentes X y Y; esto es, puede ser 1.00 en este caso y 8.75 en otro caso, lo que hace que las comparaciones de caso-con-caso sean difíciles de manejar. Necesitamos una medida que sea comparable de un problema a otro. Una medida así —y excelente— se obtiene tan sólo al escribir una fracción o tasa. Es la covarianza, CoV_{xy}, dividida entre un promedio de las varianzas de X y Y. En general, el promedio se presenta en la forma de una raíz cuadrada del producto de Vx y Vy. La fórmula completa para nuestro índice de relación entonces sería:

$$R = \frac{CoV_{xy}}{\sqrt{V_x \cdot V_y}}$$

Ésta es una forma del bien conocido coeficiente de correlación producto-momento. Al usarlo con nuestro pequeño problema obtenemos:

$$R = \frac{CoV_{xy}}{\sqrt{V_x \cdot Vy}} = \frac{1.00}{1.25} = .80$$

Este índice, que se denota por lo general como r, puede ir de +1.00 pasando por el 0 hasta −1.00, como aprendimos en el capítulo 5. Así, tenemos otra importante fuente de variación en los conjuntos de puntuaciones, siempre y cuando los elementos de los conjuntos, las X y Y, hayan sido ordenados en pares después de convertirlos en puntuaciones de desviación. Esta variación se llama de forma acertada *covarianza* y es una medida de la relación entre los conjuntos de puntuaciones.

Puede verse que la definición de relación como conjunto de pares ordenados nos lleva a varias formas de definir la relación del ejemplo anterior:

R = {(x, y); x y y son números, x siempre viene primero}
xRy = igual que arriba, o "x está relacionada con y"
R = {(0, 2),(1, 1),(2, 3),(3, 4)}
R = {(−1.5, −.5), (−.5, −1.5), (.5, .5), (1.5, 1.5)}

$$R = \frac{CoV_{xy}}{\sqrt{V_x \cdot V_y}} = \frac{1.00}{1.25} = .80$$

La varianza y la covarianza son conceptos de máxima importancia en investigación y en el análisis de los datos de investigación, por dos razones: primero, digamos que sintetizan la variabilidad de variables y la relación entre ellas. Esto se visualiza con facilidad al darnos cuenta de que las correlaciones son covarianzas que se han estandarizado para tener valores entre −1 y +1. Pero el término también implica la variación conjunta de las variables en general. En la mayor parte de nuestra investigación, literalmente perseguimos y estudiamos la covariación de los fenómenos. En segundo lugar, la varianza y covarianza forman la columna vertebral estadística del análisis multivariado, como se verá

hacia el final de este libro. La mayoría de las discusiones del análisis de datos está basada en varianzas y covarianzas. El análisis de varianza, por ejemplo, estudia diversas fuentes de variación de observaciones, en general en experimentos, como se indicó antes. El análisis factorial es en efecto el estudio de la covarianza, uno de cuyos propósitos es aislar e identificar fuentes comunes de variación. El análisis contemporáneo más reciente, el enfoque multivariado más poderoso y avanzado concebido hasta ahora, se llama *análisis de estructuras de covarianza* porque el sistema estudia conjuntos complejos de relaciones a partir del análisis de las covarianzas entre variables. Es evidente que la varianza y covarianza serán el centro de gran parte de nuestra discusión y preocupación a partir de este momento.

Anexo computacional

Uno de los mayores problemas que tienen los escritores de libros de texto en la actualidad al introducir el uso de programas de cómputo es la rapidez con que el material se vuelve obsoleto. En un periodo de un año o menos diversos fabricantes de programas estadísticos de cómputo pueden tener actualizaciones y cambios en los programas. Estas actualizaciones y modificaciones causarán una diferencia entre el programa y lo que está escrito sobre cómo usarlo. Por ejemplo, cuando la revisión de este libro de texto se inició, un programa muy popular, el paquete estadístico para ciencias sociales (SPSS) for Windows estaba en la versión 6.0. Al momento de escribir esto la versión en circulación es 8.0, y la 9.0 está por salir. Por lo anterior, presentar cualquier conjunto específico de enunciados de programación para tales programas muy pronto puede resultar inútil para los estudiantes e investigadores. Así que el objetivo es exhibir algunas características generales subyacentes a todos los programas estadísticos, que puedan tener una gran aplicación con los programas más nuevos y con los anteriores. También es importante elegir un programa estadístico que perdure por el lapso de tiempo entre las revisiones del libro (¡demasiado optimismo!). Por ejemplo, en la tercera edición de este libro publicada en 1986, las computadoras personales estaban todavía en su infancia: fuera de algunos programas de hoja de cálculo,

FIGURA 6.2

Untitled - SPSS Data Editor

File Edit View Data Transform Statistics Graphs Utilities Windows Help

	var	var	var	var	var	var	
1							
2							
3							
4							
5							
6							
7							

había un escaso desarrollo en términos de lenguajes de programación y paquetería estadística. Entre los primeros para computadoras personales o de escritorio están: SPSS-PC, STATA, NCSS y Anderson-Bell STAT. La mayor parte no eran competitivos en términos de flexibilidad y poder computacional al compararlos con el software estadístico disponible para computadoras de gran escala (macrocomputadoras). Algunos programas estadísticos especializados fueron desarrollados por un gran número de investigadores que no usaban computadoras grandes. Algunos de estos programas se escribieron en BASIC, FORTRAN, C y COBOL.

Sin embargo, el investigador actual goza la "bendición" de contar con programas estadísticos muy poderosos y flexibles que pueden usarse con facilidad para fines de análisis de datos. De aquí al final del libro, usaremos en general el SPSS for Windows para demostrar los cálculos estadísticos. ¿Por qué? En años recientes, muchas compañías competidoras de programas de cómputo (incluyendo la favorita del segundo autor) han sido adquiridas por SPSS Inc. Aún puede conseguirse paquetería de la "competencia" pero el desarrollo que conlleva es dudoso. Otra razón es que el SPSS está disponible en la mayoría de las grandes universidades y hay versiones del programa para estudiantes que pueden adquirir e instalar en sus computadoras personales. A pesar de las críticas dirigidas a SPSS, (desde su concepción) se han introducido en el mercado programas fáciles de utilizar para investigadores y estudiantes. Una vez que se clarifican algunas ideas generales en relación a cómo cargar los datos e ingresarlos en el programa, la solicitud de ciertas rutinas estadísticas se facilita mucho.

La discusión del SPSS en este libro repara sobre la versión Windows disponible para computadoras personales, y no será válida necesariamente para las versiones de macrocomputadoras o plataformas que no son Windows (tales como las versiones DOS). Esta discusión asume que el lector utiliza Windows 95 o uno más actual y que está familiarizado con los comandos y funciones de Windows. También implica el conocimiento del uso del ratón (el dispositivo apuntador) que es imperativo cuando se realizan operaciones con Windows ya que todas las operaciones se realizan al señalar con el ratón (*mouse*), haciendo clic (o resaltando) un objeto para seleccionarlo.

FIGURA 6.3

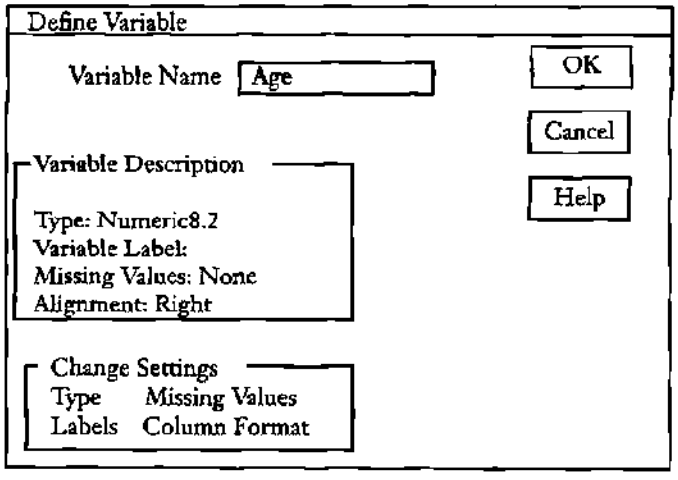


FIGURA 6.4

Untitled - SPSS Data Editor

File Edit View Data Transform Statistics Graphs Utilities Windows Help

	Age	Gender	Score	var	var	var	
1	12	2	60				
2	13	1	75				
3	15	1	45				
4	14	2	80				
5	14	1	85				
6	12	2	39				
7	13	1	62				

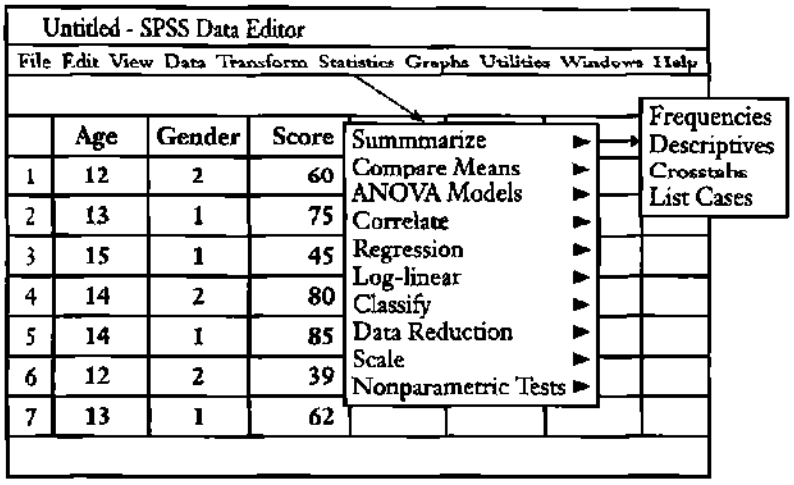
Al ejecutar el SPSS for Windows (en lo sucesivo SPSSWIN), la primera pantalla que aparece es una tabla para ingresar datos, similar a la de la figura 6.2, y tiene forma de hoja de cálculo en la que el usuario deberá incorporar los datos. Si el investigador previamente ha creado un banco de datos compatible con SPSSWIN, puede usarlo en forma directa, sin que tenga que reingresar los datos.

El formato general de datos para casi todos los programas estadísticos de computadora es una tabla donde las variables son las columnas y las observaciones (personas, individuos) son los renglones. Si tenemos el siguiente conjunto de datos, su ingreso al SPSSWIN será fácil.

	Variable		
	Año	**Género**	**Resultado de la prueba**
Personas			
1	12	M	60
2	13	F	75
3	15	F	45
4	14	M	80
5	14	F	85
6	12	M	39
7	13	F	62

Su primer paso es definir las variables para SPSSWIN. Donde usted vea la etiqueta "var" de la hoja de cálculo, usted puede ingresar el nombre para que haga referencia a sus variables. Para hacerlo, use el ratón y haga doble clic sobre la celda en la hoja de cálculo marcada con "var". Al hacerlo aparece otra pantalla que le permite especificar el nombre de la variable y sus atributos (v. gr. datos o caracteres numéricos). En la primera columna

FIGURA 6.5



escriba "Age", después haga clic en el botón "OK" (véase figura 6.3). Repita esta operación para cada una de sus variables. Para la variable "Gender" use un "1" para "F" y un "2" para "M". Después ingrese los datos de su tabla: cada valor ocupa una celda de la hoja de cálculo. Al ingresar todos los datos, su hoja de cálculo deberá ser similar a la que se muestra en la figura 6.4.

Usted puede guardarlo como un conjunto de datos al hacer clic en "FILE" y seleccionar "SAVE". Cuando lo haya hecho, se le pedirá un nombre para su conjunto de datos. Su siguiente paso es realizar un análisis estadístico. En este capítulo sólo llevaremos a cabo estadística descriptiva, que incluye medias y desviaciones estándar. La figura 6.5 muestra las pantallas usadas del SPSS. Primero haga clic en "Statistics", lo que hará aparecer un nuevo menú, del cual elija "Sumarize". Al hacerlo aparecerá un tercer menú en el cual podrá usted elegir (haciendo clic) "Descriptives", que genera una pantalla (ventana) de

FIGURA 6.6

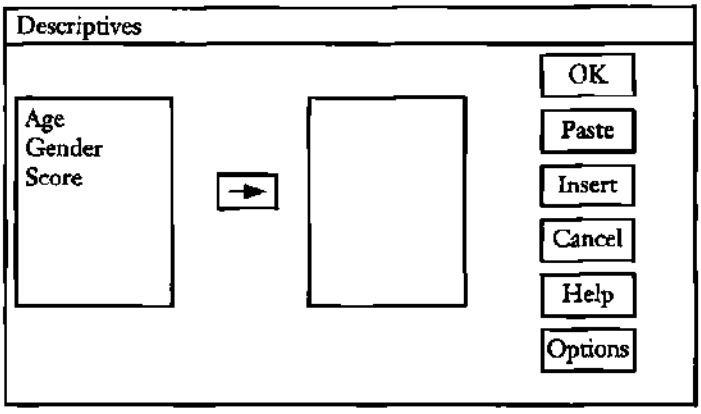
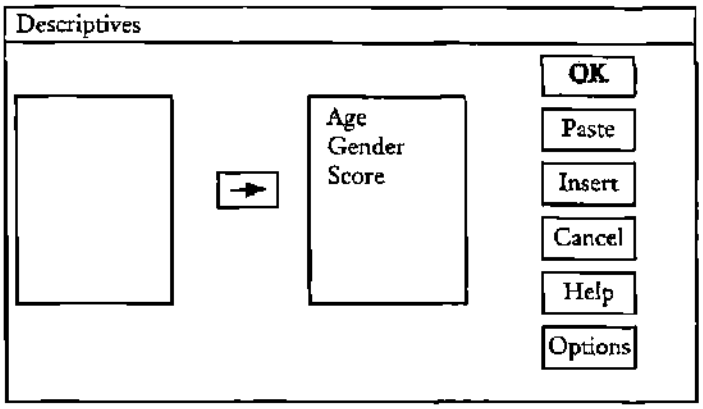


FIGURA 6.7



"estadísticos descriptivos" como se muestra en la figura 6.6. Su siguiente acción será mover las variables del cuadro de la izquierda al de la derecha. Para ello, seleccione las variables con el ratón y haga clic en el botón de la derecha. Este botón se localiza entre los dos cuadros. Se calcularán las estadísticas descriptivas sólo para aquellas que estén en el cuadro de la derecha.

Las variables que sean de poco o ningún interés para el investigador pueden quedarse en el cuadro de la izquierda. La figura 6.7 muestra la pantalla después de que las tres variables han sido desplazadas al cuadro de la derecha. El propósito de esta ventana es permitir al investigador elegir qué variables deben usarse en el análisis. Una vez realizado, seleccione el botón "OK"; entonces el programa realizará todos los cálculos necesarios y los desplegará en la pantalla de salida del SPSS.

El resultado del análisis se muestra abajo. Puede salvar los resultados en un archivo, si lo desea. Con el SPSS for Windows, puede realizar muchos análisis diferentes con el mismo conjunto de datos.

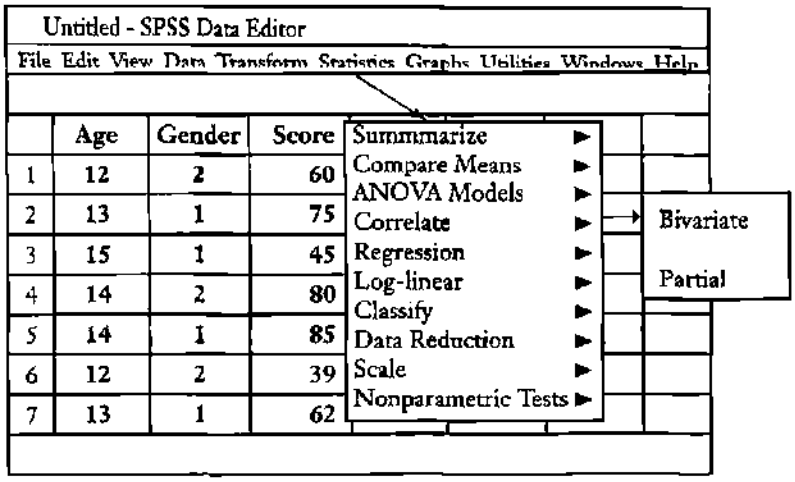
Variable	Mean	Std Dev	Variance	Minimum	Maximum	N
Gender	1.43	.53	.29	1	2	7
Age	13.29	1.11	1.24	12.00	15.00	7
Score	63.71	13.43	303.90	39.00	85.00	7

Si desea calcular las covarianzas entre variables, seleccione "Correlate" del menú, lo que generará una nueva alternativa donde puede elegir "Bivariate" (figura 6.8). Para obtener las covarianzas, seleccione el botón "options" y haga una marca en el cuadro para solicitar que se desplieguen los resultados de covarianzas.

Los resultados que da la computadora son como sigue:

Variables		Cases	Croos-Prod Dev	Variance-Covar
Age	Gender	7	−1.8571	−.3095
Age	Score	7	28.5714	4.7619
Gender	Score	7	−12.1429	−2.0238

FIGURA 6.8



En los capítulos subsiguientes cuando nos refiramos a cálculos, nos basaremos en esta demostración. La información resulta fundamental e importante para trabajar con eficiencia con el SPSS for Windows. Sin embargo, esta breve introducción no pretende ser sustituto del manual del SPSS que está disponible. El usuario de paquetería estadística debe estar consciente de que la computadora sólo calcula lo que se le pide y no puede interpretar el resultado si se han cometido algunos errores lógicos.

RESUMEN DEL CAPÍTULO

1. Las diferencias entre las mediciones son necesarias para estudiar las relaciones entre variables.
2. La varianza es una medida estadística usada para estudiar diferencias.
3. La varianza, junto con la media, se usan para resolver problemas de investigación.
4. Clases de varianza:
   a) La variabilidad de una variable o característica en el universo o población es la varianza poblacional.
   b) La muestra es un subconjunto del universo y también tiene variabilidad que se denomina varianza muestral.
   c) La estadística calculada de una muestra a la otra varía y se llama varianza muestral.
   d) La varianza sistemática es la variación de la que se puede dar cuenta. Puede ser explicada. Cualquier influencia de origen natural o humano que cause que se den eventos en alguna manera predecible es varianza sistemática.
   e) Un tipo de varianza sistemática es la llamada varianza entre grupos. Cuando hay diferencias entre grupos de sujetos, y se conoce la causa de esa diferencia, se denomina varianza entre grupos.
   f) Otro tipo de varianza sistemática es la varianza experimental, que es un poco más específica que la varianza entre grupos en tanto que está asociada con la varianza originada por la manipulación activa de la variable independiente.

g) La varianza del error es la fluctuación o variación de las medidas en la variable dependiente que no puede ser explicada en forma directa por las variables bajo estudio. Una parte de la varianza del error se debe al azar, lo que se conoce como varianza aleatoria. La fuente de esta fluctuación generalmente es desconocida. Otras posibles fuentes de varianza del error incluyen al procedimiento del estudio, el instrumento de medición y las expectativas del investigador.

5. Las varianzas pueden fraccionarse en sus componentes. En este caso, la palabra *varianza* se refiere como varianza total. La partición de la varianza total en sus componentes de varianza sistemática y varianza del error juega un papel importante en los análisis estadísticos de los datos de investigación.
6. La covarianza es la relación entre dos o más variables:
   a) Es un coeficiente de correlación no estandarizado;
   b) La covarianza y la varianza son los fundamentos estadísticos de las estadísticas multivariadas (que se presentarán en capítulos posteriores).

SUGERENCIAS DE ESTUDIO

1. Un psicólogo social ha hecho un experimento en el que un grupo, A_1, tuvo una tarea frente a un auditorio, y otro grupo, A_2, debió realizarla sin público. Las puntuaciones de ambos grupos en la tarea que evaluaba habilidad con los dedos fueron:

A_1	A_2
5	3
5	4
9	7
8	4
3	2

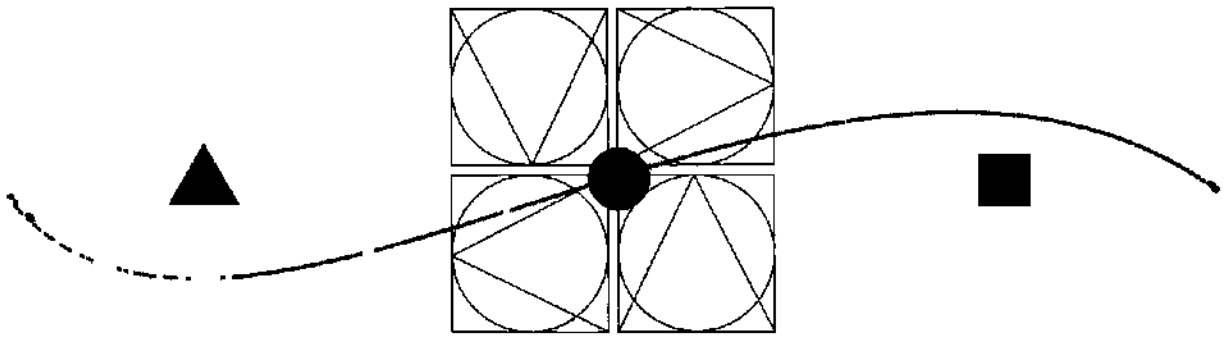
   a) Calcule las medias y varianzas de A_1 y A_2, usando el método descrito en el texto.
   b) Calcule la varianza entre grupos, V_e, y también la varianza intragrupos, V_d.
   c) Acomode todas las 10 puntuaciones en una columna y calcule la varianza total, V_t.
   d) Sustituya los valores calculados obtenidos en ***b***) y en *c*) en la ecuación: $V_t = V_e + V_d$. Interprete los resultados.
   [Respuestas: *a*) $V_{A1} = 4.8$; $V_{A2} = 2.8$; *b*) $V_e = 1.0$; $V_d = 3.8$; *c*) $V_t = 4.8$.]
2. Para el ejercicio 1 anterior, sume 2 a cada una de las puntuaciones de A_1, y calcule V_t, V_e y V_d. ¿Cuál (o cuáles) de las varianzas cambió? ¿Cuál (o cuáles) permaneció igual? ¿Por qué?
   [Respuestas: $V_t = 7.8$; $V_e = 4.0$; $V_d = 3.8$.]
3. Para el ejercicio 1 anterior, iguale las medias de A_1 y A_2, al sumar una constante de 2 a cada una de las puntuaciones de A_2. Calcule V_t, V_e y V_d. ¿Cuál es la principal diferencia entre este resultado y el de la pregunta 1? Explique por qué.
4. Suponga que un investigador social obtuvo medias de conservadurismo (A), actitud hacia la religión (B) y antisemitismo (C) de 100 individuos. Las correlaciones entre variables fueron: $r_{ab} = .70$; $r_{ac} = .40$; $r_{bc} = .30$. ¿Qué significan estas correlaciones? [Consejo: Eleve al cuadrado las *rs* antes de tratar de interpretar las relaciones. También piense en pares ordenados.]

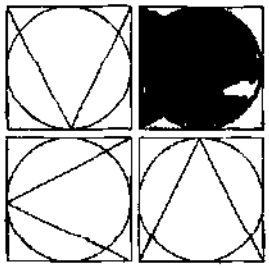
5. El propósito de esta sugerencia de estudio y de la número 6 es dotar al estudiante de intuición acerca de la variabilidad de los estadísticos muestrales, la relación entre las varianzas poblacionales y muestrales y las varianzas entre grupos y del error. El apéndice C contiene 40 conjuntos de 100 números aleatorios entre 0 y 100, con sus medias, varianzas y desviaciones estándar. Extraiga 10 conjuntos de 10 números cada uno a partir de 10 lugares diferentes en la tabla.
   a) Calcule la media, varianza y desviación estándar de cada uno de los 10 conjuntos. Encuentre la media más alta y la más baja, así como la varianza más alta y la más baja. ¿Difieren mucho una de la otra? ¿Qué valor "deberían" tener las medias (50)? Una vez hecho esto, aparte los 10 totales y calcule la media de los 100 números. ¿Difieren mucho las 10 medias de la media total? ¿Difieren mucho de las medias reportadas en la tabla de medias, varianzas y desviaciones estándar que se presenta después de los números aleatorios?
   b) Cuente los números pares y nones en cada uno de los 10 conjuntos. ¿Son los que "deberían ser"? Cuente los números pares e impares de los 100 números. ¿El resultado es "mejor" que los resultados de los 10 conteos? ¿Por qué debería serlo?
   c) Calcule la varianza de las 10 medias. Ésta es, por supuesto, la varianza entre grupos V_b. Calcule la varianza del error usando la fórmula: $V_e = V_t - V_b$.
   d) Analice el significado de sus resultados después de repasar la explicación en el texto.
6. Tan pronto como sea posible, los estudiantes de investigación deben empezar a entender y utilizar la computadora. La sugerencia de estudio 5 se puede realizar mejor y más fácilmente con la computadora. Sería mejor, por ejemplo, extraer 20 muestras de 100 números cada una. ¿Por qué? En cualquier caso, los estudiantes deben aprender cómo realizar operaciones estadísticas simples utilizando el equipo de cómputo y los programas que existen en sus instituciones. Todas las instituciones poseen programas de cómputo para calcular medias y desviaciones estándar (las varianzas se obtienen elevando al cuadrado las desviaciones estándar)[4] y para generar números aleatorios. Si usted tiene acceso al equipo de su institución, utilícelo para llevar a cabo la sugerencia de estudio 5, pero incremente el número de muestras y sus n.

[4] Puede hacer pequeñas discrepancias entre las desviaciones estándar y las varianzas calculadas a mano y aquellas obtenidas a partir de la computadora, porque los programas existentes y las rutinas de las calculadoras manuales generalmente usan una fórmula con N menos 1 en lugar de colocar N en el denominador de la fórmula. Sin embargo, las discrepancias serán pequeñas, sobre todo si la N es grande. (La razón para que existan diferentes fórmulas se explicará más adelante cuando nos ocupemos de muestreo y otros temas.)

Parte Tres

Probabilidad y muestreo





Capítulo 7

Probabilidad

- Definición de probabilidad
- Espacio muestral, puntos muestrales y eventos
- Determinación de probabilidades con monedas
- Un experimento con dados
- Algo de teoría formal
- Eventos compuestos y su probabilidad
- Independencia, exclusión mutua y exhaustividad
- Probabilidad condicional
  - **Definición de probabilidad condicional**
  - **Un ejemplo académico**
  - **Teorema de Bayes: revisión de las probabilidades**

El tema de la probabilidad es simple y obvio, confuso y complejo. Es un tema sobre el que sabemos mucho y a la vez es un tema del cual no sabemos nada. Los alumnos de jardín de niños y los filósofos pueden estudiar probabilidad. Es tedioso y es interesante. Tales contradicciones son características de la probabilidad.

Si se utiliza la expresión "leyes del azar", dicha expresión, por sí misma, es particularmente contradictoria. Casualidad o azar, por definición, es la ausencia de ley. Si los eventos pueden explicarse por medio de una ley, entonces no se deben al azar, por lo tanto, ¿por qué decimos "leyes del azar"? La respuesta es también peculiarmente contradictoria: es posible obtener conocimiento de la ignorancia si se ve el azar como ignorancia. Esto es debido a que los eventos aleatorios, en conjunto, ocurren obedeciendo leyes con regularidad monótona. A partir del desorden del azar, el científico unifica el orden de la predicción y el control científicos.

No es fácil explicar estos conceptos desconcertantes, de hecho, los filósofos discrepan en sus respuestas. Afortunadamente, no hay desacuerdo sobre los eventos probabilísticos empíricos —o al menos hay muy poco—. Casi todos los científicos y filósofos concordarían en que si dos dados se lanzan un número de veces, probablemente habrá más 7 que 2 o 12.

También estarían de acuerdo en que ciertos eventos, como encontrarse un billete de 100 dólares o ganar una combinación de apuestas, son extremadamente improbables.

Definición de probabilidad

¿Qué es la probabilidad? Al hacer esta pregunta surge inmediatamente un problema complejo. Wang (1993), Brady y Lee (1989) y Cowles (1989) han establecido que históricamente parece no haber acuerdo sobre la respuesta. Esto puede deberse a que existen dos escuelas principales de pensamiento: la de frecuencia y la de no frecuencia. Además, en la escuela de frecuencia hay por lo menos dos definiciones, entre otras, que parecen ser irreconciliables: la *a priori* y la *a posteriori*. La definición *a priori* se debe al controvertido Pierre Laplace y al distinguido matemático Augustus DeMorgan (Cowles, 1989). Aquí, la probabilidad de un evento es igual al número de casos favorables dividido entre el número total de casos (igualmente posibles), o $p = f \div (f + u)$, donde p es la probabilidad, f es el número de casos favorables y u el número de casos no favorables. El método para calcular la probabilidad, implicado en la definición, es *a priori* en el sentido de que la probabilidad está dada de tal manera que se pueden determinar las probabilidades de eventos antes de la investigación empírica. Las personas frecuentemente hacen afirmaciones respecto de las probabilidades sin datos empíricos que las avalen. Aunque estas afirmaciones reflejan, más bien, un punto de vista propio. La interpretación de la probabilidad de Laplace y DeMorgan se considera una definición clásica, la cual es la base de la probabilidad teórica matemática.

La definición *a posteriori*, o de frecuencia relativa a largo plazo, es empírica por naturaleza. Ésta explica que, en una serie real de pruebas, la probabilidad es la razón del número de veces en que un evento ocurre en el número total de ensayos. Con esta definición, uno se aproxima a la probabilidad de forma empírica aplicando una serie de pruebas, contando el número de veces en que un cierto evento ocurre y después calculando la relación. El resultado del cálculo es la probabilidad de dicha clase de evento. Tienen que usarse definiciones de frecuencia cuando no es posible la enumeración teórica sobre clases de eventos. Por ejemplo, para calcular la probabilidad de la longevidad y la que tiene un caballo en una carrera, se tienen que usar tablas actuariales y calcular la probabilidad a partir de conteos y cálculos pasados. La afirmación de que un cortador de diamantes es 95% preciso, indica que de cada cien diamantes que esta persona ha cortado en el pasado, 95 de ellos fueron cortados correctamente.

Hablando prácticamente (y para los propósitos del texto), la distinción entre la definición *a priori* y la *a posteriori* no es demasiado importante. Siguiendo a Margenau (1950/1977, p. 264), unimos las dos definiciones diciendo que el enfoque *a priori* provee una definición constitutiva de probabilidad, mientras que el enfoque *a posteriori* ofrece una definición operacional de probabilidad. Se requiere usar ambos enfoques, ya que se necesita complementar una con la otra.

El planteamiento de la no frecuencia es atribuida a John Maynard Keynes (1921/1979). Keynes, un economista de fama mundial, escribió un número importante de publicaciones citadas frecuentemente. Existe una teoría económica completa que está basada en las contribuciones de Keynes. Aquellos que trabajan en tal clase de investigación son llamados *keynesianos*. La contribución de Keynes a la probabilidad y la estadística generalmente no se menciona en la mayoría de los libros de texto de probabilidad, pero aún así es importante para aquellos que hacen investigación en las ciencias del comportamiento (Brady y Lee, 1989a). En este enfoque hay dos valores: 1) el valor de la probabilidad en sí misma y 2) el peso de una evidencia asociada a ella. El peso de la evidencia es subjetiva, involucra

la percepción de quien toma la decisión sobre la cualidad y cantidad de información alrededor del valor de la probabilidad, obtenido empíricamente. En esencia Keynes establece que quienes toman las decisiones se confrontan con la probabilidad de los eventos, y también con la cantidad y/o cualidad de la información asociada a ellos. Quienes toman las decisiones utilizan la información aunada a la probabilidad para tomar una decisión. Keynes define un coeficiente de peso y riesgo; dicho coeficiente asigna, esencialmente, un peso a un valor empírico de probabilidad. Si el peso de la evidencia es fuerte, a la probabilidad se le da un mayor peso. Se asigna un peso cercano a cero si el peso de la evidencia respecto de esa probabilidad es débil. Según Brady y Lee (1989b, 1991) el enfoque de Keynes explica algunas de las llamadas paradojas de la toma de decisiones, que la teoría de frecuencia no puede explicar adecuadamente. Bakan (1974) afirma que la teoría de la probabilidad de Keynes capta la esencia del proceso que enfrentan los psicólogos clínicos que tratan con problemas de relevancia. Al llevar a cabo una terapia, el psicólogo escucha, lee y ve muchos indicios e información, pero selectivamente ubica a algunos de ellos como más relevantes que otros. La teoría de Keynes posee un mayor alcance en las explicaciones probabilísticas y puede utilizarse para explicar el resultado de un estudio de Rosenthal y Gaito, reportado en Bakan (1974), donde se le pidió a un grupo de profesores psicólogos doctorados juzgar dos estudios diferentes: *A* y *B*. En cada uno de los estudios *A* y *B* se había realizado la misma prueba estadística y se había obtenido el mismo valor de p. Sin embargo, el tamaño de la muestra del estudio *A* era de 10 y el del estudio *B* era de 100. A cada profesor de la facultad se le preguntó cuál de los estudios le inspiraba mayor confianza o credibilidad. La mayoría de ellos le otorgó mayor confianza a los resultados del estudio *B*. Keynes explicaría esto a la luz del hecho de que en su juicio, estos individuos daban mayor peso a un tamaño muestral de 100 que a un tamaño muestral de 10.

En resumen, la frecuencia relativa a largo plazo es la teoría prevalente en la investigación de las ciencias del comportamiento (Cowles, 1989). La mayoría de los científicos del comportamiento confían la manipulación estadística de sus datos, siguen la escuela de la frecuencia relativa del pensamiento. En casi todos los textos elementales de estadística que cubren el tema de la probabilidad, solamente se discute la teoría de la frecuencia relativa y sus efectos en los métodos estadísticos.

Espacio muestral, puntos muestrales y eventos

Para calcular la probabilidad de cualquier resultado primero es necesario determinar el número total de resultados posibles. En un dado, los resultados posibles son 1, 2, 3, 4, 5, 6. Llamemos a este conjunto *U*, ya que es el espacio muestral o universo de posibles resultados. El espacio muestral incluye todos los posibles resultados en un "experimento" que son de interés para el experimentador. Los elementos primarios de *U* son llamados elementos o puntos muestrales. Se escribe, entonces, $U = \{1, 2, 3, 4, 5, 6\}$, para unificar este capítulo con el razonamiento y método de conjuntos empleados en los capítulos 4, 5 y 6. Si x_j es igual a cualquier punto muestral o elemento en *U*, se escribe ahora $U = \{x_1, x_2, \ldots, x_n\}$. Todos los posibles resultados de lanzar dos dados son ejemplos de diferentes *U* (véase tabla 7.1); todos los niños de un jardín de niños en tal o cual sistema escolar; todos los votantes elegibles en *X* país.

Algunas veces la determinación de un espacio muestral es fácil, pero algunas veces es difícil; el problema es análogo a la definición de conjuntos del capítulo 4. Los conjuntos pueden definirse listando a todos lo miembros del conjunto y estableciendo una regla para la inclusión de los elementos en él. En la teoría de la probabilidad, ambas definiciones se utilizan. ¿Qué valor tendría *U* al lanzar al aire 2 monedas? He aquí una lista de todas las

TABLA 7.1 *Matriz de posibles resultados con dos dados.*

		Segundo dado					
		1	2	3	4	5	6
	1	2	3	4	5	6	7
	2	3	4	5	6	7	8
Primer	3	4	5	6	7	8	9
dado	4	5	6	7	8	9	10
	5	6	7	8	9	10	11
	6	7	8	9	10	11	12

posibilidades: $U = \{(C, C), (C, X), (X, C), (X, X)\}$.* Ésta es una definición listada de U; una definición por regla —aunque no la usaremos— podría ser $U = \{x\}$; donde x representa todas las combinaciones de C y X. En este caso U es un producto cartesiano. Que sea $A_1 = \{C_1, X_1\}$, el resultado de la primera moneda y $A_2 = \{C_2, X_2\}$, el de la segunda moneda. Recordando que un producto cartesiano de dos conjuntos es el conjunto de todos los pares ordenados, cuya primera entrada es un elemento de un conjunto y la segunda entrada un elemento de otro conjunto, podemos esquematizar la generación del producto cartesiano de este caso, $A_1 \times A_2$, como en la figura 7.1. Note que hay cuatro líneas conectando a A_1 y a A_2, por lo que hay cuatro posibilidades: $\{(C_1, C_2), (C_1, X_2), (X_1, C_2), (X_1, X_2)\}$. Este esquema de pensamiento y procedimiento puede utilizarse para definir muchos espacios muestrales de *Us*, aunque el procedimiento real puede ser tedioso.

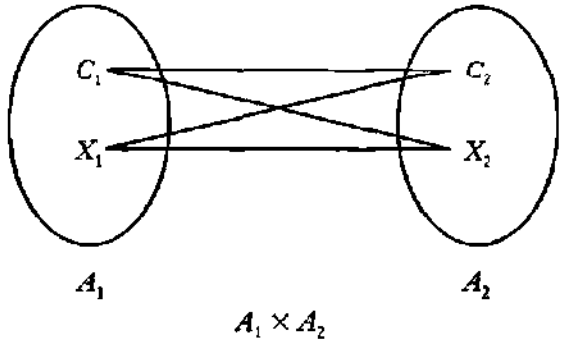
En caso de usar dos dados, ¿qué sería U? Si se piensa en el producto cartesiano de dos conjuntos, probablemente surja un poco de problema. Suponga que A_1 son los resultados o los puntos del primer dado: {1, 2, 3, 4, 5, 6} y que A_2 son los resultados o los puntos del segundo dado, entonces $U = A_1 \times A_2 = \{(1, 1), (1, 2), \ldots, (5, 6), (6, 6)\}$. Esto se puede ilustrar como se hizo con el ejemplo de las monedas, pero contar las líneas es más difícil debido a que habrá demasiadas. Se puede conocer el número de posibles resultados simplemente realizando la operación $6 \times 6 = 36$, o en una fórmula: mn, donde m es el número de posibles resultados del primer conjunto y n es el número de posibles resultados del segundo conjunto.

A menudo es posible resolver problemas difíciles de probabilidad usando diagramas de árbol. Éstos definen espacios muestrales y posibilidades lógicas con claridad y precisión. Un árbol es un diagrama que da todas las posibles alternativas o resultados en las combinaciones de conjuntos, al proporcionar rutas y puntos del conjunto. Esta definición es un poco difícil de manejar por lo que se requiere ilustrarla: tomemos el ejemplo de las monedas (se coloca el árbol sobre un costado). El diagrama de árbol se muestra en la figura 7.2.

Para determinar el número de posibles alternativas se cuenta el número de alternativas o puntos ubicados en la "cima" del árbol. En este caso hay cuatro alternativas y para nombrarlas se leen, para cada punto final, los puntos que llevan a él. Por ejemplo, la primera alternativa es (C_1, C_2). Obviamente tres, cuatro o más monedas pueden ser usadas, el único problema es que el procedimiento es tedioso por el gran número de alternativas. El diagrama para tres monedas se ilustra en la figura 7.3; aquí existen ocho posibles alternativas, resultados o puntos muestrales: $U = \{(C_1, C_2, C_3), (C_1, C_2, X_3), \ldots, (X_1, X_2, X_3)\}$ (los elementos de este conjunto se llaman *tríos ordenados*).

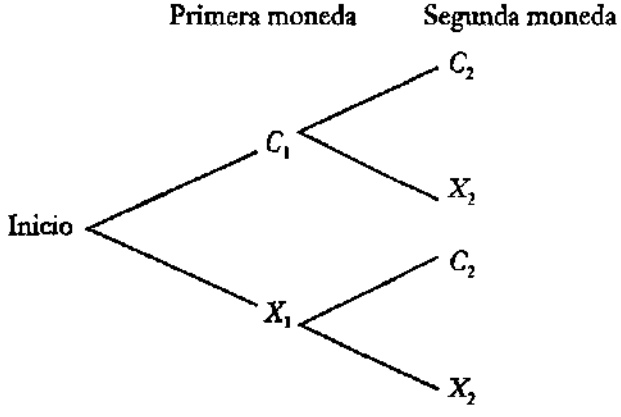
* C y X corresponden a los nombres dados a los lados de una moneda, "cara" y "cruz". En lo sucesivo se utilizará C para referirse a "cara" y X para referirse a "cruz". (*N. del T.*)

FIGURA 7.1



Los puntos muestrales de un espacio muestral pueden parecer un poco confusos al lector porque se han discutido dos clases de puntos sin haber hecho alguna diferenciación. Esta confusión puede aclararse por medio del uso de un término: Un evento es un subconjunto de U; cualquier elemento de un conjunto es también un subconjunto del conjunto. Recuerde que con el conjunto $A = \{a_1, a_2\}$ por ejemplo, ambos $\{a_1\}$ y $\{a_2\}$ son subconjuntos de A, así como lo son $\{a_1, a_2\}$ y $\{\ \}$, es el conjunto vacío. Igualmente, todos los resultados de las figuras 7.2 y 7.3, por ejemplo, (C_1, X_2), (X_1, C_2) y (X_1, C_2, X_3) son subconjuntos de sus respectivos U, por lo tanto, por definición también son eventos. Pero en el uso cotidiano, un evento abarca un poco más que los puntos. Todos los puntos son eventos (subconjuntos), pero no todos los eventos son puntos, es decir, un punto o resultado es una clase especial de evento: la clase más simple. Siempre que se establece una proposición, se describe un evento. Por ejemplo: "si se lanzan dos monedas al aire, ¿cuál es la probabilidad de tener dos caras?" Las "dos caras" es un evento, que si sucede, en este caso, es también un punto muestral. Pero si la pregunta fuera: "¿Cuál es la probabilidad de obtener al menos una cara?". "Al menos una cara" es un evento, pero no un punto muestral porque incluye, en este caso, tres puntos muestrales: (C_1, C_2), (C_1, X_2) y (X_1, C_2) (véase figura 7.2).

FIGURA 7.2



Determinación de probabilidades con monedas

Si se lanza una moneda recién acuñada 3 veces y se anota $p(C) = 1/2$ y $p(X) = 1/2$, que significa que la probabilidad de caras es 1/2 y similar para las cruces, se supone, entonces, equiprobabilidad. El espacio muestral para los tres lanzamientos de monedas (o un lanzamiento de tres monedas) es: $U = \{(C, C, C), (C, C, X), (C, X, C), (C, X, X), (X, C, C), (X, C, X), (X, X, C), (X, X, X)\}$. Note que si no se presta atención al orden de caras y cruces, se obtiene un caso de tres caras, un caso de tres cruces, tres casos de dos caras y una cruz, y tres casos de dos cruces y una cara. La probabilidad de cada uno de los ocho resultados es obviamente 1/8, la probabilidad de tres caras es 1/8 y la de tres cruces es 1/8. La probabilidad de dos caras y una cruz es, por otro lado, 3/8 y de forma similar para la probabilidad de dos cruces y una cara.

Las probabilidades de todos los puntos en el espacio muestral deben sumar 1.00, y las probabilidades siempre son positivas. Si se realiza un diagrama de árbol de probabilidad para el experimento de los tres lanzamientos de moneda, se parecerá al de la figura 7.3. Cada rama completa del árbol (desde el inicio hasta el tercer lanzamiento) es un punto muestral y todas las ramas comprenden el espacio muestral. Las secciones de cada rama (o ruta) están etiquetadas con la probabilidad; en este caso todas son etiquetadas con 1/8. Esto conduce naturalmente al enunciado de un principio básico: si los resultados en diferentes puntos del árbol (en el primero, segundo y tercer lanzamiento) son independientes uno de otro (esto es, si un resultado no influye a otro en forma alguna), entonces la probabilidad de cualquier punto muestral (*CCC*, quizá) es el producto de las probabilidades de los resultados aislados. Por ejemplo la probabilidad de tres caras es $1/2 \times 1/2 \times 1/2 = 1/8$.

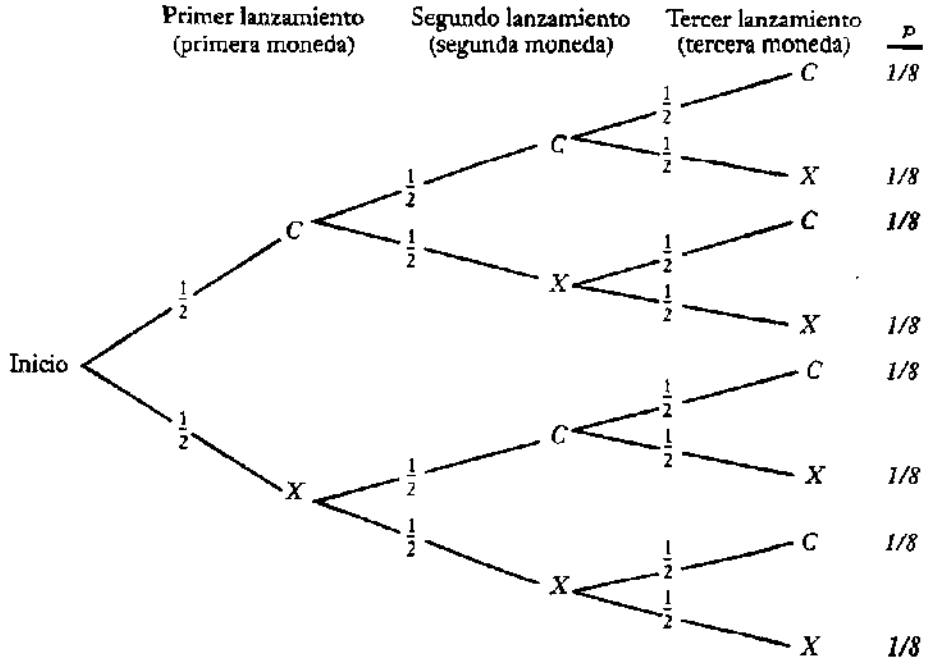
Otro principio implica que para obtener la probabilidad de cualquier evento se sumen las probabilidades de los puntos muestrales que componen dicho evento. Por ejemplo, ¿cuál es la probabilidad de obtener dos caras y una cruz? Si se buscan en el árbol las ramas que tienen dos caras y una cruz existen tres de ellas, que se observan en la figura 7.3. Así, $1/8 + 1/8 + 1/8 = 3/8$. En el lenguaje de conjuntos, se encuentran los subconjuntos (eventos) de U y se observan sus probabilidades. Los subconjuntos de U del tipo "2 caras y 1 cruz" son, partiendo del árbol o de la definición previa de U: $\{(C, C, X), (C, X, C), (X, C, C)\}$. Si se llama A a este conjunto o evento, entonces $p(A_1) = 3/8$.

Este procedimiento puede seguirse con un experimento laborioso de cien lanzamientos de moneda al aire, pero en su lugar, para obtener las expectativas teóricas, solamente se multiplica el número de lanzamientos por la probabilidad de cualquiera de ellas para tener el número esperado de caras (o cruces). Esto puede hacerse porque *todas* las probabilidades son iguales. Una pregunta importante que debe hacerse ahora es: en experimentos reales en los que cien monedas fueran lanzadas al aire, ¿obtendríamos exactamente 50 caras, suponiendo que las monedas están niveladas? No, no muy a menudo, serían aproximadamente 8 veces en 100 repeticiones de esta naturaleza. Esto puede escribirse $p = 8/100$, o 0.08. (Las probabilidades pueden escribirse en forma decimal o de fracción, aunque es más frecuente usar la forma decimal.)

Un experimento con dados

Si lanzamos dos dados recién fabricados 72 veces bajo condiciones cuidadosamente controladas; si sumamos el número de puntos en los dos dados en los 72 lanzamientos, obtendremos un conjunto de sumas que van de 2 a 12. Algunos de estos resultados (sumas) serán más frecuentes que otros, simplemente porque hay más formas de obtener ciertos resultados. Por ejemplo, solamente hay una forma de obtener 2 o 12: 1 + 1 y 6 + 6, pero

FIGURA 7.3



existen 3 maneras de obtener 4: 1 + 3, 3 + 1 y 2 + 2. Si esto es cierto, entonces las probabilidades para obtener diferentes sumas deben ser diferentes. El juego de los dados está basado en estas diferencias en las frecuencias esperadas.

Para resolver un problema de probabilidad a priori, primero se debe definir el espacio muestral: U = {(1, 1), (1, 2), (1, 3), ..., (6, 4), (6, 5), (6, 6)}; es decir, se aparea cada número del primer dado con cada número del segundo dado en turno (otra vez el producto cartesiano). Esto puede verse fácilmente si se coloca este procedimiento en una matriz (véase tabla 7.1). Suponiendo que queremos conocer la probabilidad del evento "obtener 7", simplemente contamos el número de 7 en la tabla. Se encuentran seis de ellos a lo largo de la diagonal central. Existen 36 puntos muestrales en *U*, obtenidos por medio de enumerarlos, como se hizo anteriormente, o usando la fórmula mn. Esta fórmula implica multiplicar el número de posibilidades del primer evento por el número de posibilidades del segundo evento. Este método puede definirse de la siguiente forma: suponga que hay m formas de hacer algo, *A*, y que hay n formas de hacer otra cosa, *B*; si las n formas de hacer *B* son independientes de las m formas de hacer *A*, entonces hay $m \times n$ formas de hacer ambas, *A* y *B*. Este principio puede extenderse para más de dos cosas. Si, por ejemplo, hay tres cosas *A*, *B* y *C*, donde hay r formas de hacer *C*, entonces la fórmula es mnr.

Si se aplica esta fórmula al problema de los dados, entonces $mn = 6 \times 6 = 36$, y suponiendo equiposibilidad otra vez, la probabilidad de cualquier resultado simple es 1/36. La probabilidad de obtener un 12, por ejemplo, es de 1/36. Sin embargo, la probabilidad de obtener un 4 es diferente, dado que el 4 ocurre tres veces en el cuadro anterior. Se deben sumar las probabilidades para cada uno de estos elementos del espacio muestral: 1/36 + 1/36 + 1/36 = 3/36; así $p(4) = 3/36 = 1/12$. Como se vio anteriormente, la probabilidad de

un 7 es $p(7) = 6/36 = 1/6$; la probabilidad de un 8 es $p(8) = 5/36$. Note también que podemos calcular las probabilidades de combinaciones de eventos. Los jugadores frecuentemente apuestan a tales combinaciones. Por ejemplo, ¿cuál es la probabilidad de obtener un 4, o un 10? En lenguaje de conjuntos ésta es una pregunta de unión: $p(4 \cup 10)$. Se cuenta el número de 4 y de 10 en el cuadro; hay tres 4 y tres 10. Así $p(4 \cup 10) = 6/36$.

En la tabla 7.1, por medio del conteo de las probabilidades de cada tipo de resultado se puede extraer una tabla de frecuencias esperadas (f_e) para 36 lanzamientos; después se duplican estas frecuencias para obtener las frecuencias esperadas (*a priori*) para 72 lanzamientos; se confrontan las frecuencias esperadas (f_e) contra las frecuencias obtenidas (f_o) cuando dos dados son tirados realmente 72 veces, y finalmente aparecen las diferencias absolutas entre las frecuencias esperadas y las frecuencias obtenidas (los resultados se encuentran en la tabla 7.2). Las discrepancias no son grandes, de hecho, con una prueba estadística no difieren significativamente de lo esperado por efectos del azar. El método *a priori* parece tener sus virtudes.

Algo de teoría formal

Se tiene un espacio muestral U, con los subconjuntos A, B,... Los elementos de U (y de A, B,...) son a_1, b_1,... esto es $a_1, a_2, \ldots, a_n$ y $b_1, b_2, \ldots b_n$, etcétera. A, B y los demás son eventos. En realidad, aunque se ha hablado de la probabilidad de una ocurrencia aislada, de hecho se refiere a la probabilidad de un tipo de ocurrencia. Se puede hablar acerca de la probabilidad de cualquier evento aislado de U, por ejemplo, porque cualquier miembro particular de U es concebido como representativo de todo U, y también para las probabilidades de los subconjuntos A, B, ..., K de U. La probabilidad de U es 1; la probabilidad de E (el conjunto vacío) es 0 (cero); o $p(U) = 1.00$; $p(E) = 0$. Para determinar la probabilidad de cualquier subconjunto de U, debe asignarse una medida del conjunto. Para asignar tal medida, se da un peso a cada elemento de U y también a cada elemento de los subconjuntos de U. Un *peso* es definido por Kemeny, Snell y Thompson (1974) como sigue:

> Un peso es un número positivo asignado a cada elemento, x, en U, y escrito $w(x)$, de tal manera que la suma de todos estos pesos, $\Sigma w(x)$, es igual a 1.

Ésta es una noción de función; w es llamada una función del peso. Es una regla que asigna pesos a los elementos de un conjunto U, en forma tal que la suma de los pesos es igual a 1; esto es, $w_1 + w_2 + w_3 + \ldots + w_n = 1.00$, y $w_i = 1/n$. Los pesos son iguales, asumiendo equiprobabilidad; cada peso es una fracción con 1 en el numerador y el número de casos n en el denominador. En el experimento previo de lanzar una moneda (figura 7.3), los pesos asignados a cada elemento de U, siendo U *todos* los resultados, son igual a 1/8. La suma de todas las funciones del peso $w(x)$, es $1/8 + 1/8 + \ldots + 1/8 = 1$. En la teoría de la probabilidad, la suma de los elementos del espacio muestral debe ser siempre igual a 1.

TABLA 7.2 *Frecuencias esperadas y obtenidas de las sumas de dos dados lanzados 72 veces*

Suma de los dados	2	3	4	5	6	7	8	9	10	11	12
$f_e(36)$	1	2	3	4	5	6	5	4	3	2	1
$f_e(72)$	2	4	6	8	10	12	10	8	6	4	2
$f_o(72)$	4	2	6	6	10	15	7	11	6	4	1
Diferencia	2	2	0	2	0	3	3	3	0	0	1

Obtener la medida de un conjunto a partir de los pesos es fácil: la medida de un conjunto es la suma de los pesos de los elementos de dicho conjunto.

$$\sum_{x \text{ en } U} w(x) \quad \text{o} \quad \sum_{x \text{ en } A} w(x)$$

(Observe que la suma de los pesos en un subconjunto A, de U, no tiene que ser igual a 1. De hecho, usualmente es menor que 1.)

Escribir $m(A)$, significa "La medida del conjunto A". Esto simplemente habla de la suma de los pesos de los elementos del conjunto A.

Si se selecciona una muestra al azar, a partir de 400 alumnos de cuarto grado de un sistema escolar, entonces U son todos los 400 alumnos. Cada alumno es un punto muestral de U. Cada alumno es una x en U. La probabilidad de seleccionar cualquier niño al azar es 1/400. Suponga que A es igual a los niños en U, y B es igual a las niñas en U, y que hay 100 niños y 300 niñas. A cada niño le es asignado un peso de 1/400 y a cada niña un peso de 1/400. Si se desea una muestra de 100 alumnos en total, la expectativa es, entonces, 25 niños y 75 niñas en la muestra. La medida del conjunto A, $m(A)$, es la suma de los pesos de todos los elementos en A. Dado que hay 100 niños en U, sumamos los 100 pesos: 1/400 + 1/400 + ... + 1/400 = 100/400 = 1/4 o:

$$m(A) = \sum_{x \text{ en } A} w(x) = \frac{1}{4}$$

De manera similar:

$$m(B) = \sum_{x \text{ en } B} w(x) = \frac{3}{4}$$

Para el conjunto B (las niñas), se suman 300 pesos, siendo cada uno de ellos de 1/400. En pocas palabras, la suma de los pesos representa las probabilidades, es decir que la medida de un conjunto es la probabilidad de que un miembro del conjunto sea elegido. Así, se puede decir que la probabilidad de que un miembro de la muestra de 400 alumnos sea niño es de 1/4, y la probabilidad de que el miembro seleccionado sea niña es de 3/4. Para determinar las frecuencias esperadas, se multiplica el tamaño de la muestra por estas probabilidades: $1/4 \times 100 = 25$ y $3/4 \times 100 = 75$.

La probabilidad tiene tres propiedades fundamentales:

1. La medida de cualquier conjunto, como se definió anteriormente, es mayor o igual a 0 y menor o igual a 1, es decir, las probabilidades (medidas de conjuntos) pueden ser 0, 1 o alguna cantidad entre ellos.
2. La medida de un conjunto $m(A)$, es igual a 0, si y sólo si no hay miembros en A; esto es, si A está vacío.
3. Suponga que A y B son los conjuntos. Si A y B están separados, esto es, $A \cap B = E$, entonces $m(A \cup B) = m(A) + m(B)$.

Esta ecuación dice que cuando A y B no tienen miembros en común, entonces tanto la probabilidad de A como de B, o de ambos es igual a las probabilidades combinadas de A y B.

No hay necesidad de un ejemplo para ilustrar el punto 1), ya se han dado varios anteriormente. Para ilustrar 2), suponga que en el ejemplo de niños y niñas queremos saber cuál es la probabilidad de tener en la muestra a un maestro, pero U no incluye maestros. Si

C es el conjunto de los maestros de cuarto grado, en este caso, dicho conjunto está vacío y $m(C) = 0$. Con el mismo ejemplo de alumnos niños y niñas se ilustra 3): si A es el conjunto de niños y B el conjunto de niñas, entonces $m(A \cup B) = m(A) + m(B)$ pero $m(A \cup B) = 1.00$ porque ellos son los únicos subconjuntos de U. Debido a que $m(A) = 1/4$ y que $m(B) = 3/4$, la ecuación se sostiene.

Eventos compuestos y su probabilidad

Anteriormente se dijo que un evento es un subconjunto de U, pero es necesario detallar esto. Un evento es un conjunto de posibilidades: es un conjunto de eventos posibles; es el resultado de un "experimento" de probabilidad. Un evento compuesto es la co-ocurrencia de dos o más eventos aislados (o compuestos). Las dos operaciones de conjuntos de intersección y unión —las operaciones que más interesan aquí— implican eventos compuestos. Si se lanza una moneda y un dado, el resultado es un evento compuesto y se puede calcular la probabilidad de tal evento. Aún más interesante, se puede preguntar cómo están relacionadas ciertas variables demográficas. Una forma de hacer esto es buscar respuestas a preguntas tales como: "¿Cuál es la probabilidad de detectar a un usuario de drogas que elige días específicos para usar la droga, sin considerar ninguna estrategia de prueba de drogas?" (véase Borack, 1997) o, "¿Cuál es la probabilidad de que dos estudiantes en el mismo salón de clases tengan el mismo mes y día de nacimiento?"(véase Nunnikhoven, 1992), o "¿Cuál es la probabilidad de que un estudiante de posgrado abandone sus estudios?"(véase Cooke, Sims & Peyrefitte, 1995).

Los eventos compuestos son más interesantes que los eventos aislados —y más útiles en investigación—. Con ellos pueden estudiarse las relaciones. Para entender esto, primero proceda a definir e ilustrar qué son los eventos compuestos y después a examinar ciertos problemas de conteo y las formas en que el conteo está relacionado con la teoría de conjuntos y la teoría de la probabilidad. Se encontrará que si la teoría básica es entendida, la aplicación de la teoría de probabilidad a los problemas de investigación se facilita considerablemente. Además, la interpretación de los datos se ve menos sujeta a errores.

Suponga que se ha estudiado un grupo de niños de una escuela primaria y que hay 100 niños en total en dicho grupo: 60 de cuarto grado y 40 de sexto grado. La función numérica es útil ya que asigna a cualquier conjunto el número de miembros en el conjunto. El número de miembros en A es $n(A)$. En este caso $n(U) = 100$, $n(A) = 60$, y $n(B) = 40$, donde A es el conjunto de los alumnos de cuarto grado y B es el conjunto de los alumnos de sexto grado, ambos subconjuntos de U, que representa los 100 alumnos de la escuela primaria. Si no hay traslape entre ambos conjuntos, $A \cap B = E$, entonces la siguiente ecuación se sostiene:

$$n(A \cup B) = n(A) + n(B) \tag{7.1}$$

Recuerde que anteriormente la definición de frecuencia de probabilidad fue dada como sigue:

$$p = \frac{f}{f + u} \tag{7.2}$$

donde f es el número de casos favorables y u el número de casos desfavorables. El numerador es $n(f)$ y el denominador es $n(U)$, el número total de casos posibles. De manera similar se pueden dividir los términos de la ecuación 7.1 entre $n(U)$:

$$\frac{n(A \cup B)}{n(U)} = \frac{n(A)}{n(U)} + \frac{n(B)}{n(U)} \tag{7.3}$$

Esto se reduce a probabilidades análogamente a la ecuación 7.2:

$$p(A \cup B) = p(A) + p(B) \tag{7.4}$$

Usando el ejemplo de los 100 alumnos de escuela y sustituyendo los valores en la ecuación 7.3, se obtiene:

$$\frac{100}{100} = \frac{60}{100} + \frac{40}{100},$$

lo que se sustituye por la ecuación 7.4:

$$1.00 = .60 + .40$$

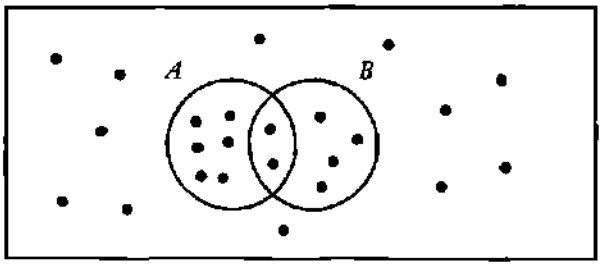
En muchos casos, dos (o más) conjuntos en los que nos interesamos, no están separados sino que de hecho se traslapan. Cuando esto sucede, entonces $A \cap B \neq E$ y no es cierto que $n(A \cup B) = n(A) + n(B)$. Observe la figura 7.4, en ésta A y B son subconjuntos de U; los puntos muestrales están indicados por puntos. El número de puntos muestrales en A es 8; el número de puntos muestrales en B es 6. Hay dos puntos muestrales en $A \cap B$. por lo que la ecuación anterior no se sostiene. Si se calculan todos los puntos en $A \cup B$ con la ecuación 7.1, se obtiene $8 + 6 = 14$ puntos; pero existen solamente 12 puntos, por lo que esta ecuación debe modificarse en una ecuación más general que abarque todos los casos:

$$n(A \cup B) = n(A) + n(B) - n(A \cap B) \tag{7.5}$$

Debe quedar claro que cuando se usa la ecuación 7.1, el error resulta de contar los dos puntos de $A \cap B$ dos veces. Por lo tanto, se substrae una vez $n(A \cap B)$, para corregir la ecuación. Ahora se ajusta a cualquier posibilidad. Si, por ejemplo, $n(A \cap B) = E$, el conjunto vacío, la ecuación 7.5 se reduce a la ecuación 7.1. La ecuación 7.1 es un caso especial de la ecuación 7.5. Si se calcula el número de puntos muestrales en $n(A \cap B)$ de la ecuación 7.4, entonces se obtiene $n(A \cup B) = 8 + 6 - 2 = 12$. Si se divide la ecuación 7.5 entre $n(U)$, como en la ecuación 7.3, resulta:

$$p(A \cup B) = p(A) + p(B) - p(A \cap B) \tag{7.6}$$

FIGURA 7.4



Si se sustituyen las marcas o puntos muestrales, se encuentra que:

$$\frac{12}{24} = \frac{8}{24} + \frac{6}{24} - \frac{2}{24}$$

$$.50 = .33 + .25 - .08$$

Entonces en una muestra aleatoria de U, las probabilidades de que un elemento sea miembro de A, B, $A \cap B$ y $A \cup B$, son respectivamente: .33, .25, .08 y .50.

Independencia, exclusión mutua y exhaustividad

Considere las siguientes preguntas, variantes de las que deben hacerse los investigadores. ¿La ocurrencia de este evento A impide la posibilidad de ocurrencia de este otro evento B? ¿La ocurrencia del evento A tiene alguna influencia en la ocurrencia del evento B? ¿Están relacionados los eventos A, B y C? ¿Cuando A ocurre, esto tiene influencia en los resultados de B y quizás C? ¿Los eventos A, B, C y D agotan las posibilidades? ¿O hay quizás, otras posibilidades E, F, etcétera? Por ejemplo, un investigador está estudiando las decisiones de un comité de educación y su relación con la preferencia política, preferencia religiosa, educación y otras variables. Para relacionar estas variables con las decisiones del comité, el investigador ha de tener algún método para clasificar las decisiones. Una de las primeras preguntas que deben hacerse es "¿Mi sistema de clasificación agota todas las posibilidades?" Una pregunta adicional es "Si el comité toma una clase de decisión, ¿excluirá esto la posibilidad de que se tome otra decisión?" Quizás la pregunta más importante que el investigador puede hacer es "¿Si el comité toma una decisión particular, influirá ésta en cualquier otra decisión?"

Se ha hablado de exhaustividad, exclusión mutua e independencia. Ahora se definen estas ideas de una manera más detallada y se usan en ejemplos de probabilidad. Su aplicabilidad general e importancia se harán evidentes en los capítulos donde se estudia el análisis de datos.

Suponga que A y B son subconjuntos de U. ¿Hay otros subconjuntos de U (además del conjunto vacío)? ¿Agotan A y B el espacio muestral? ¿Están todos los puntos muestrales del espacio muestral U incluidos en A y B? Un ejemplo simple es: Establezca que $A = \{C, X\}$; y que $B = \{1, 2, 3, 4, 5, 6\}$. Si se lanzan simultáneamente una moneda y un dado, ¿cuáles son las posibilidades resultantes? A menos que todas las posibilidades estén agotadas, no se puede resolver el problema de probabilidad. Hay 12 posibilidades (2×6). Los conjuntos A y B agotan el espacio muestral (esto, por supuesto, es obvio dado que A y B generan el espacio muestral). Ahora tomemos un ejemplo más real: un investigador está estudiando las preferencias religiosas y utiliza el siguiente sistema para categorizar a los individuos: protestantes, católicos y judíos. Implícitamente, U es establecido de tal forma que incluya a toda la gente (con o sin preferencia religiosa) y a todos los subconjuntos de U: A es igual a protestantes, B es igual a católicos y C es igual a judíos. La pregunta de conjuntos es: ¿$A \cup B \cup C = U$? ¿Se han agotado todas las preferencias religiosas? ¿Que hay de los budistas? ¿Y de los musulmanes? ¿Y de los ateos?

La *exhaustividad*, entonces significa que los subconjuntos de U agotan todo el espacio muestral, o que $A \cup B \cup \ldots \cup K = U$, donde A, B, ..., K son subconjuntos de U, el espacio muestral. En lenguaje de probabilidad, esto significa: $p(A \cup B \cup \ldots \cup K) = 1.00$. A menos que el espacio muestral U, haya sido agotado, por así decirlo, las probabilidades no pueden calcularse adecuadamente. Por ejemplo, en el caso de la preferencia religiosa,

suponga que $A \cup B \cup C = U$ pero que de hecho hubiera un gran número de individuos sin una preferencia religiosa en particular. Así que en realidad, $A \cup B \cup C \cup D = U$, donde D es el subconjunto de individuos sin una preferencia religiosa. Las probabilidades calculadas en el supuesto de esta ecuación serían muy diferentes de aquellas basadas en el supuesto de la ecuación previa.

Dos eventos, A y B, son *mutuamente excluyentes* cuando no están unidos, o cuando $A \cap B = E$, es decir, cuando la intersección de dos (o más) conjuntos es el conjunto vacío (o cuando dos conjuntos no tienen elementos en común), se dice que ambos son mutuamente excluyentes. Esto es igual a decir, otra vez en lenguaje de probabilidad, $p(A \cap B) = 0$. Es más conveniente para los investigadores que los eventos sean mutuamente excluyentes, porque entonces pueden sumar las probabilidades de los eventos. Un principio, en términos de conjuntos y probabilidades es el siguiente: si los eventos (conjuntos) A, B y C son mutuamente excluyentes, entonces $p(A \cup B \cup C) = p(A) + p(B) + p(C)$. Éste es el caso especial de un principio más general que se discutió en una sección previa (véase ecuaciones 7.1, 7.4, 7.5 y 7.6 y la argumentación que las explica).

Uno de los principales propósitos del diseño de investigación es establecer las condiciones de independencia de los eventos, de tal manera que las condiciones de dependencia de éstos puedan ser estudiadas adecuadamente. Dos eventos A y B, son estadísticamente independientes si se ajustan a la ecuación:

$$p(A \cap B) = p(A) \cdot p(B) \tag{7.7}$$

que dice que la probabilidad de que A y B ocurran, es igual a la probabilidad de A por la probabilidad de B. Ejemplos fáciles y claros de eventos independientes son el lanzamiento de monedas y dados. Si A es el evento de lanzar un dado y B es el evento de lanzar una moneda al aire, y $p(A) = 1/6$ y $p(B) = 1/2$, entonces, si $p(A) \cdot p(B) = 1/6 \cdot 1/2 = 1/12$, entonces A y B son independientes. Si lanzamos una moneda 10 veces, un lanzamiento no tiene influencia en ningún otro lanzamiento; los lanzamientos son independientes y sucede lo mismo cuando se lanza un dado. De manera similar, cuando lanzamos una moneda al aire y al mismo tiempo se lanza un dado, los eventos de lanzar el dado, A, y lanzar la moneda, B, son independientes. El resultado de lanzar el dado no tiene influencia en el lanzamiento de la moneda, y viceversa. Desafortunadamente, este claro modelo no siempre se puede aplicar a situaciones de investigación.

La creencia, basada en el sentido común, de la llamada ley de promedios es tremendamente errónea, pero ilustra el poco entendimiento que existe respecto del concepto de independencia. Esta dice que si hay un gran número de ocurrencias de un evento, la posibilidad de que ese evento ocurra en el siguiente ensayo es mínima. Supongamos que una moneda es lanzada al aire, y que resultan caras en cinco veces seguidas. La "ley de los promedios" haría creer que existe una mayor posibilidad de obtener cruz en la siguiente lanzada, pero no es así. La probabilidad sigue siendo 1/2. Cada lanzamiento es un evento independiente.

Suponga que los estudiantes de una clase universitaria están tomando un examen, y que lo están realizando bajo las condiciones usuales de no comunicación, no ver el documento del compañero, etcétera. Las respuestas de cada uno pueden ser consideradas independientes de las respuestas de cualquier otro estudiante. ¿Pueden considerarse independientes las respuestas a los reactivos dentro de cada examen? Suponga que la respuesta a un reactivo posterior en el examen está insertada en reactivo previo en el mismo examen. Digamos que la probabilidad de tener correcto el reactivo posterior debido al azar es 1/4, pero el hecho de que la respuesta fuese dada antes puede cambiar esta probabilidad. Para algunos estudiantes puede llegar a ser de 1.00. Lo que es importante para el

investigador es conocer que esa independencia es frecuentemente difícil de lograr y que la carencia de independencia, cuando la investigación supone que la hay, puede afectar seriamente la interpretación de los datos.

Suponga que se ordenan por rangos los exámenes y luego se les asignan calificaciones con base en dicho orden. Éste es un procedimiento perfectamente legítimo y útil, pero debe reconocerse que las calificaciones dadas con el método de ordenar por rangos no son independientes (si acaso pudieran serlo). Si se toman cinco de esos exámenes y después de leerlos, uno de ellos, es ordenado en primer lugar (como el mejor), el siguiente se ordena como segundo, y así con los cinco exámenes. Se le asigna el número "1" al primero, el número "2" al segundo, el número "3" al tercero, el número "4" al cuarto y el número "5" al quinto. Después de usar el número 1, sólo quedan, 2, 3, 4 y 5. Después del número 2, quedan solamente el 3, 4 y 5. Cuando se asigna el 4, obviamente debe asignarse el 5 al examen restante; en pocas palabras, la asignación del 5 se ve influida por la asignación del 4 (y también por las del 1, 2 y 3). La asignación de los eventos no fue independiente. Aquí podría surgir la pregunta, "¿Y esto, importa?". Suponga que tomamos los rangos, y se los usamos como puntuaciones, para luego hacer inferencias acerca de las diferencias de las medias entre los grupos, digamos, entre dos clases. La prueba estadística usada para hacer esto está probablemente basada en el paradigma de la moneda y los dados, con su independencia original, pero aquí no se ha seguido este modelo (uno de sus más importantes supuestos, la independencia, ha sido ignorada).

Cuando investigamos eventos que carecen de independencia, las pruebas estadísticas carecen de cierta validez. Una prueba de χ^2, por ejemplo, supone que los eventos (respuestas de los individuos a una pregunta de entrevista) registrados en las casillas de una tabla de contingencia, son independientes una de otra. Si los eventos registrados no son independientes uno de otro, entonces las bases de la prueba estadística y las inferencias hechas a partir de ésta, han sido alteradas.

Considere la investigación de las relaciones entre la autoeficacia percibida y el comportamiento. En este ejemplo, los investigadores intentan mostrar la congruencia entre el juicio de la autoeficacia y el comportamiento real. A los participantes generalmente se les aplican escalas de autoeficacia que describen un conjunto de tareas bien definidas. A cada participante se le pide juzgar si puede completar cada tarea. La percepción y el comportamiento son congruentes cuando la percepción del sujeto iguala a la ejecución real, esto es, "dijo que puede hacerlo y lo hizo" y "dijo que no puede hacerlo y en realidad no lo hizo". Cervone (1987) afirma que un gran número de investigadores que trabajan en esta área han reportado congruencias excepcionalmente altas. Numerosos estudios han encontrado congruencias del 80-90%. Cervone señala, sin embargo, que los datos obtenidos de las escalas de autoeficacia no son independientes ya que cada individuo participante contribuye con más de una observación al análisis. Cervone (1987, p.710) afirma que "como en cualquier área de investigación, uno no debe suponer que las observaciones múltiples de un mismo sujeto sean independientes".

Kramer y Schmidhammer (1992) encontraron problemas de independencia similares en investigaciones de conducta animal. Algunos estudios sobre conducta humana y conducta animal dependen de datos etológicos de frecuencia. Este tipo de datos generalmente cuentan el número de encuentros entre organismos (animales o humanos) o de la ejecución de un comportamiento. Kramer y Schmidhammer usan el ejemplo de la medición del comportamiento de la rana. Para obtener observaciones independientes del comportamiento de vocalización o no vocalización de la rana macho a lo largo de una sección de la orilla de un lago, los investigadores necesitan comprobar si la ausencia o presencia de la vocalización de una rana no tiene efecto en otras ranas. Kramer y Schmidhammer observaron que muchos patrones de comportamiento de interés para el etólogo tienden a

ocurrir en conglomerados y no en forma independiente. Kramer y Schmidhammer citan muchos estudios que pueden tener un problema potencial de independencia.

Otro estudio es el realizado por Keane (1990), en el cual examinó la preferencia de ratas macho de patas blancas por hembras en etapa de estro. Keane registró el número de encuentros que cada rata hembra en etapa de estro tenía con cada ratón macho; dichos encuentros fueron clasificados como agresivos (pelea o persecución) o amistosos (acicalar u olfatear). El origen de la rata macho fue documentado a partir del momento del nacimiento, para que el experimentador supiera cómo cada rata hembra estaba emparentada con cada macho. Keane deseaba saber si la hembra en etapa de estro prefería a un macho emparentado o a uno que no lo fuera. Keane encontró que la rata hembra mostraba una conducta más amistosa y menos conductas agresivas hacia sus primos hermanos que hacia los machos con quienes no estaba emparentada. Si se toman en consideración los puntos del artículo de Kramer y Schmidhammer, el estudio de Keane sería imperfecto en el hecho de que sus observaciones pueden no ser independientes. Una rata hembra pudo haber tenido un especial "disgusto" hacia un no pariente o un "gusto" especial por un primo hermano, y los conteos de frecuencia podrían estar inclinados a favor de ese par de machos.

Ahora consideremos un estudio antiguo, pero importante, sobre el comportamiento agresivo de los simios, realizado por Hebb y Thompson en 1968. Los datos de su investigación se presentan en la tabla 7.3. El problema trataba acerca de la relación entre sexo y agresión. Se tomaron muestras del comportamiento de 30 chimpancés adultos con el fin de estudiar diferencias individuales en el temperamento de los simios. Sin entrar en detalles, puede decirse que un análisis de las observaciones mostró que tanto los machos como las hembras presentaron un comportamiento amistoso casi con igual frecuencia, pero los machos eran más agresivos. Los resultados de Hebb y Thompson parecerían decir: "¡cuidado con los machos!", pero los autores señalaron que esto difiere de la experiencia que tienen los cuidadores de simios. ¡19 de 20 arañazos y cortaduras fueron infligidos por hembras! Entonces Hebb y Thompson buscaron la interesante, aunque desconcertante idea de tabular la incidencia de actos agresivos de dos maneras: cuando éstos eran precedidos por una cuasi–agresión, es decir, por una alerta de ataque, y cuando los actos agresivos eran precedidos por una conducta amistosa. Las incidencias resultantes del comportamiento buscado parecen indicar: ¡cuidado con las hembras cuando son amistosas!". Los machos fueron los únicos en mostrar actos cuasi–agresivos antes de actos verdaderamente agresivos (37 actos de machos y 0 actos de hembras), sin embargo, sólo las hembras actuaron en forma agresiva después de mostrar una conducta amistosa (15 actos de hembra y 0 actos de machos).

Estos datos no pueden ser analizados estadísticamente en forma válida dado que los números indican la frecuencia de tipos de actos, así, todos los 37 actos de los machos podrían haber sido realizados sólo por uno o dos de ellos. Si un simio hubiera cometido todos los 37 actos, entonces debería quedar claro que los actos no eran independientes uno de otro; dicho simio podría tener mal carácter y las conductas por mal carácter notoriamente carecen de independencia en los actos humanos y animales.

El siguiente ejemplo es hipotético: un investigador realiza un muestreo de 100 decisiones de un comité de educación. Hay una gran variedad de formas de hacer esto. Muchas decisiones pueden ser muestreadas de pocos comités, o muchas decisiones pueden ser muestreadas de muchos comités, o ambos. Si el investigador desea asegurarse de la independencia de las decisiones, entonces la mayoría de las decisiones deberán ser muestreadas de muchos comités de educación. Teóricamente, sólo una decisión debería ser tomada de cada comité. Esto nos da alguna seguridad de independencia —al menos tanta como sea posible—. Tan pronto como más de una decisión es tomada de un mismo comité, el inves-

tigador debe considerar el hecho de que las decisiones de clase A pueden influir en las decisiones de clase B. La decisión A puede influir en la decisión B, por ejemplo, porque los miembros del comité deseen parecer consistentes. Ambas decisiones pueden involucrar gastos de equipo instruccional, y si el comité adopta una política liberal respecto de A, entonces deberá adoptar una política similar con B.

Suponga que un investigador ha calculado la probabilidad de la diferencia entre dos medias, en donde dicha diferencia fue debida al azar. La probabilidad fue de 5/100, o 0.05, lo que indica que hubo aproximadamente 5 oportunidades en 100 de que el resultado obtenido fuera debido al azar, es decir, que si la condición experimental es repetida 100 veces sin manipulación experimental, aproximadamente 5 de estas 100 veces podría dar una diferencia de medios tan grande como la obtenida con la manipulación experimental. Sintiéndose dudoso acerca del resultado —después de todo existen 5 oportunidades en 100 de que el resultado pueda ser debido al azar— el investigador repitió cuidadosamente el experimento completo, y se obtuvo el mismo resultado (¡suerte!). Habiendo controlado todo cuidadosamente para asegurarse de que los dos experimentos fueran independientes, la probabilidad calculada para los dos resultados fue debida al azar. Esta probabilidad fue aproximadamente de 0.02. Así se pueden observar tanto los valores de independencia en el experimento como la importancia de la replicación de los resultados.[1]

Hay que hacer notar, finalmente, que la fórmula para la independencia trabaja en dos sentidos. 1) indica la probabilidad de que ambos eventos ocurran al azar, si los eventos son independientes y se conocen las probabilidades de los eventos por separado. Si se encontrara que los dados repetidamente muestran, digamos 12, entonces probablemente algo ande mal con los dados. Si un jugador observa que otro jugador parece ganar siempre, el jugador que va perdiendo, por supuesto, tendrá sospechas. Las posibilidades de ganar continuamente un juego limpio son pocas; puede suceder, por supuesto, pero es muy poco probable que así sea. En investigación es poco probable que se obtengan dos o tres resultados significativos por azar; en ese caso es probable que algo más allá del azar esté operando (la variable independiente, se espera). 2) La fórmula para independencia puede invertirse, por así decirlo; puede decírsele a los investigadores qué hacer para sacar ventaja de las probabilidades multiplicativas. El investigador debe, si esto es posible, planear la investigación para que los eventos sean independientes, aunque esto es más fácil decirlo que hacerlo, lo cual se hará evidente antes de finalizar este libro.

Probabilidad condicional

En toda investigación —y quizá especialmente en la investigación científica social y educacional— los eventos a menudo no son independientes. Se puede ver la independencia de otra manera. Cuando dos variables están relacionadas, éstas no son independientes. La discusión previa sobre conjuntos aclara: si $A \cap B = E$, entonces no hay relación (específicamente una relación cero), o A y B son independientes. Si $A \cap B \neq E$, entonces hay una relación, o A y B no son independientes. Cuando los eventos no son independientes,

[1] El método para calcular estas probabilidades combinadas fue propuesto por Fisher y se describe en Mosteller y Bush (1954). El estudiante perspicaz puede preguntarse por qué el principio de conjuntos aplicado a la probabilidad, $p(A \cap B) = p(A) \cdot p(B)$, no es aplicable, es decir, ¿por qué no calcular $.05 \times .05 = 0.0025$? Mosteller y Bush explican este punto, pero dado que éste es punto difícil y discutible no se incluye en el presente texto. Todo lo que el lector necesita hacer es recordar que la probabilidad de obtener una diferencia sustancial entre las medias en la misma dirección en experimentos repetidos es considerablemente más pequeña que tener tal diferencia una sola vez. Así, uno puede estar más seguro de los datos y conclusiones propias que los de otros datos que resulten iguales.

los científicos pueden afinar sus inferencias probabilísticas. El significado de esta afirmación puede explicarse, hasta cierto punto, al estudiar la probabilidad condicional.

Cuando los eventos no son independientes, el enfoque de la probabilidad debe ser modificado. Un ejemplo simple es el siguiente: ¿cuál es la probabilidad de que de cualquier pareja casada, tomada al azar, ambos miembros sean republicanos? Primero, suponiendo que existe equiprobabilidad y que todo lo demás es igual, el espacio muestral U (todas las posibilidades) es $\{RR, RD, DR, DD\}$, donde la esposa aparece primero en cada posibilidad o punto muestral; de este modo la probabilidad de que ambos, esposo y esposa sean republicanos es $p\{RR\} = 1/4$. Pero si ya sabe que uno de ellos es republicano, ¿cuál es la probabilidad de que ambos sean republicanos ahora? U se reduce a $\{RR, RD, DR\}$. El conocimiento de que uno de ellos es republicano elimina la posibilidad DD, y de esta forma se reduce el espacio muestral, por lo tanto, $p(RR) = 1/3$. En caso de tener, además, la información de que la esposa es republicana, ¿cuál es la probabilidad de que ambos esposos sean republicanos? Ahora $U = \{RR, RD\}$, y por lo tanto $p(RR) = 1/2$. Las nuevas probabilidades son, en este caso "condicionadas" por un conocimiento previo de los hechos.

Definición de probabilidad condicional

Suponga que A y B son los eventos en el espacio muestral, U, como lo hemos estado haciendo. La probabilidad condicional se expresa: $p(A \mid B)$, que se lee: "la probabilidad de A, dado B". Por ejemplo, se podría decir: "La probabilidad de que un esposo y su esposa sean, ambos, republicanos, siendo que el esposo es republicano," o, mucho más difícil de contestar, aunque más interesante: "la probabilidad de una labor docente universitaria altamente efectiva, dado el grado académico de doctor". Por supuesto, puede escribirse también $p(B \mid A)$. La fórmula para la probabilidad condicional cuando se involucran dos eventos es:

$$p(A \mid B) = \frac{p(A \cap B)}{p(B)} \tag{7.8}$$

La fórmula toma una noción anterior de probabilidad y la altera por las situaciones de probabilidad condicional. (Note por favor que la teoría de la probabilidad condicional se extiende a más de dos eventos, pero esto no se analizará en este libro.) Es importante recordar que en problemas de probabilidad el denominador debe ser el espacio muestral. La fórmula anterior cambia el denominador de la razón y por lo tanto cambia el espacio muestral, el cual ha sido reducido de U a B. Para demostrar este punto tomemos dos ejemplos: uno de independencia o probabilidad simple y otro de dependencia o probabilidad condicional.

Si lanzamos al aire una moneda dos veces, los eventos son independientes. ¿Cuál es la probabilidad de obtener cara en el segundo lanzamiento, si la cara apareció en el primero?

TABLA 7.3 *Matriz de probabilidad que muestra las probabilidades conjuntas de dos eventos independientes.*

		Segundo lanzamiento		
		Cara_2	Cruz_2	
Primer lanzamiento	Cara_1	1/4	1/4	1/2
	Cruz_1	1/4	1/4	1/2
		1/2	1/2	

TABLA 7.4 *Matriz de probabilidades conjuntas de eventos.*

		Segundo lanzamiento		
		$Cara_2$	$Cruz_2$	
Primer lanzamiento	$Cara_1$	.30	.20	.50
	$Cruz_1$	.30	.20	.50
		.60	.40	1.00

Esto ya se sabe: 1/2. Si se calcula la probabilidad usando la ecuación 7.8, primero se hace una matriz de probabilidad (véase la tabla 7.3). Para las probabilidades de cara (C) y cruz (X) en el primer lanzamiento, se leen las entradas marginales en el lado derecho de la matriz; igualmente para las probabilidades del segundo lanzamiento, que están en la parte inferior de la matriz. De este modo $p(C_1) = 1/2$, $p(C_2) = 1/2$, y $p(C_1 \cap C_2) = 1/4$. Por lo tanto:

$$p(C_2 \mid C_1) = \frac{p(C_2 \cap C_1)}{p(C_1)} = \frac{\frac{1}{4}}{\frac{1}{2}} = \frac{1}{2}$$

Los resultados concuerdan con el razonamiento simple previo. Sin embargo, si se hace el problema un poco más complejo, quizás la fórmula sea más útil. Suponga que la probabilidad de obtener cara en el segundo lanzamiento fuera .60 en lugar de .50, y que los eventos continúan siendo independientes. ¿Cambia esto la situación? Esta nueva situación se ilustra en la tabla 7.4 (El .30 en la celda $C_1 \cap C_2$ se calcula con las probabilidades en los márgenes: .50 × .60 = .30. Esto es permitido dado que sabemos que los eventos son independientes. Si no fueran independientes, los problemas de probabilidad condicional no podrían resolverse sin el conocimiento de por lo menos uno de los valores.) La fórmula nos da:

$$p(C_2 \mid C_1) = \frac{p(C_2 \cap C_1)}{p(C_1)} = \frac{.30}{.50} = .60$$

Pero este .60 es el mismo que la probabilidad simple de $Cara_2$. Cuando los eventos son independientes, se obtienen los mismos resultados. Esto es, en este caso: $p(C_2 \mid C_1) = p(C_2)$ y en el caso general:

$$p(A \mid B) = p(A) \tag{7.9}$$

Se obtiene otra definición o condición de independencia. Si se mantiene la ecuación 7.9, los eventos son independientes.

Un ejemplo académico

Hay más ejemplos interesantes de probabilidad condicional que los de monedas y otros mecanismos del azar. Tomemos el problema desconcertante y frustrante de predecir el éxito de los estudiantes de doctorado en una universidad. ¿Podría ser usado el modelo de moneda y dados en una situación tan compleja? Sí, bajo ciertas condiciones. Desafortunadamente estas condiciones son difíciles de ajustar, aunque ha habido éxito limitado. En caso de tener cierta información empírica, el modelo puede ser muy útil. Suponga que los administradores de una universidad están interesados en predecir el éxito de sus estudiantes de doctorado, ya que se encuentran preocupados por el pobre rendimiento de

TABLA 7.5 *Probabilidades conjuntas, problema de los graduados universitarios.*

		Éxito (E)	Fracaso (F)	
MAT	≥65	.20	.10	.30
MAT	<65	.20	.50	.70
		.40	.60	1.00

muchos de ellos y desean establecer un sistema de selección. La universidad continúa admitiendo a todos los aspirantes al doctorado, como en el pasado, pero por tres años todos los estudiantes aceptados presentaron el Miller Analogies Test (MAT), un examen que ha probado ser útil al predecir el éxito en la universidad en muchas áreas (por ejemplo, psicología, educación, economía). Esta prueba también ha sido usada para evaluar personal para puestos de alto nivel en la industria. Se selecciona un punto de corte arbitrario en el puntaje bruto de 65.

La administración universitaria encuentra que el 30% de todos los candidatos del periodo de tres años tuvo una puntuación de 65 o más. Cada uno se categoriza como éxito (E) o fracaso (F). El criterio es simple: ¿Obtuvo el grado el estudiante? Si lo obtuvo, entonces es definido como éxito. Se encuentra que 40% del número total fueron exitosos. Para determinar la relación entre la puntuación MAT y el éxito o fracaso, la administración, empleando nuevamente el punto de corte de 65, determina las proporciones mostradas en la tabla 7.5. El MAT divide al grupo exitoso en dos (.20 y .20) pero diferencia marcadamente en el grupo de fracaso (.10 y .50). Ahora las preguntas que se hacen son: ¿cuál es la probabilidad de obtener el grado de doctorado si el candidato recibe una puntuación de 65 o mayor en el MAT? ¿Cuál es la probabilidad de que un candidato obtenga el grado si la puntuación MAT es menor de 65? Los cálculos son

$$p(E \mid \geq 65) = \frac{p(E \cap \geq 65)}{p(\geq 65)} = \frac{.20}{.30} = .67$$

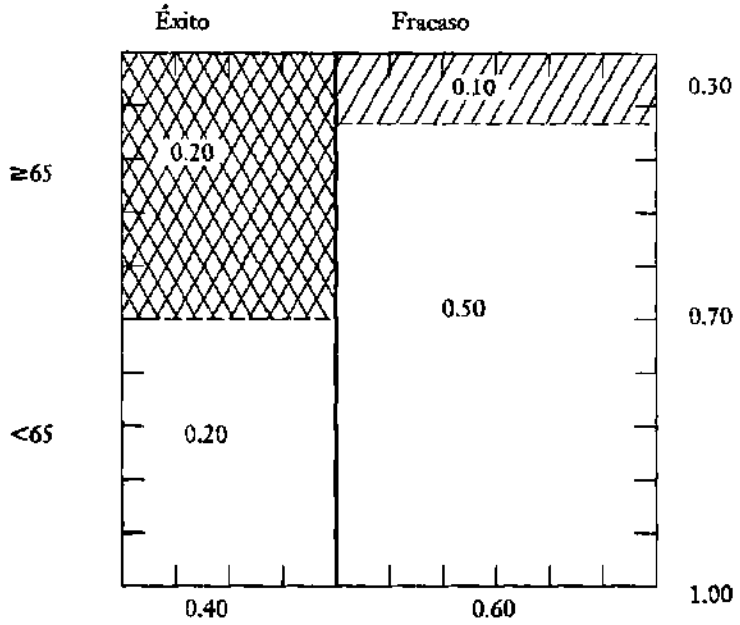
$$p(E \mid < 65) = \frac{p(E \cap < 65)}{p(< 65)} = \frac{.20}{.70} = .29$$

Claramente puede verse que el MAT es un buen predictor de éxito en el programa.

Note cuidadosamente lo que sucede en todos estos casos: cuando escribimos $p(A \mid B)$ en lugar de la sola $p(A)$ en efecto reducimos el espacio muestral de U a B. Si se toma el ejemplo anterior, la probabilidad de éxito sin ningún otro conocimiento es un problema de probabilidad en todo el espacio muestral U. Esta probabilidad es de .40, pero conociendo la puntuación MAT, el espacio muestral se reduce de U a un subconjunto $U \geq 65$. El número real de ocurrencias de un evento exitoso, por supuesto, no cambia; el mismo número de personas tiene éxito, pero la fracción de probabilidad tiene un nuevo denominador; en otras palabras, el estimado de la probabilidad es refinado por el conocimiento de subconjuntos "pertinentes" de U. En este caso, ≥ 65 y < 65 son subconjuntos "pertinentes" de U. Al decir subconjuntos "pertinentes" se quiere decir que la variable implicada está relacionada con la variable criterio: éxito y fracaso.

La siguiente forma de ver el problema puede ayudar. Una interpretación por áreas, del problema de los estudiantes graduados, ha sido representada en la figura 7.5. Aquí se emplea la idea de una medida de un conjunto. Recuerde que una medida de un conjunto o subconjunto es la suma de los pesos de dicho conjunto o subconjunto, y que tales pesos

FIGURA 7.5



son asignados a los elementos del conjunto o subconjunto. La figura 7.5 es un cuadrado con 10 partes iguales en cada lado y cada parte es igual a 1/10 o .10. El área de todo el cuadrado es el espacio muestral U, y la medida de U, $m(U)$, es igual a 1.00. Esto significa que todos los pesos asignados a todos los elementos del cuadrado suman en total 1.00. Las medidas de los subconjuntos han sido insertadas: $m(F) = .60$, $m(<65) = .70$, $m(S \cap \geq 65) = 0.20$. Las medidas de estos subconjuntos pueden ser calculadas multiplicando las longitudes de sus lados. Por ejemplo, el área de la caja superior izquierda (con doble trazado diagonal) es de $.5 \times .4 = .20$. Recuerde que la probabilidad de cualquier conjunto o subconjunto es la medida de dicho conjunto (o subconjunto). Así que la probabilidad de cualquiera de las cajas en la figura 7.5 es la que se indica en cada una da ellas. Se puede calcular la probabilidad de cualesquiera de las dos cajas sumando las medidas de los conjuntos; por ejemplo, la probabilidad de éxito es .20 + .20 = .40.

Estas medidas (o probabilidades) están definidas en el área total, o $U = 1.00$. La probabilidad de éxito es igual a .40/1.00. Como se conoce el rendimiento de los estudiantes en el MAT, las áreas que indican las probabilidades asociadas con ≥65 y <65 están delimitadas por los guiones horizontales. La probabilidad simple de ≥65 es igual a .20 + .10 = .30, o .30/1.00. Toda el área sombreada de la parte superior indica esta probabilidad, mientras que las áreas de las medidas de éxito y fracaso están indicadas por las líneas gruesas que las separan en el cuadrado.

El problema de probabilidad condicional es el siguiente: ¿Cuál es la probabilidad de éxito, dado el conocimiento de la puntuación del MAT, o dado ≥65 (también podría ser <65 por supuesto)? Se tiene un pequeño nuevo espacio muestral, indicado por toda el área sombreada en la parte superior del cuadrado; U ha sido reducido a este espacio más pequeño porque se conoce la "verdad" del pequeño espacio. En lugar de dejar que este

pequeño espacio sea igual a .30, ahora supone que es igual a 1.00. (Podría parecer que tenemos un nuevo *U*). En consecuencia, las medidas de las cajas que constituyen el nuevo espacio muestral deben ser recalculadas. Por ejemplo, en lugar de calcular la probabilidad de $p(\geq 65 \cap S) = .20$ porque es de 2/10 del área total del cuadrado, se debe calcular la probabilidad con base en el área de ≥65 (el área sombreada en la parte superior del cuadrado), debido a que se sabe que los elementos en el conjunto ≥65 tienen puntajes mayores o iguales a 65. Una vez hecho esto, se obtiene .20/.30 = .67 [el área sombreada para éxito ÷ el área sombreada para éxito + el área sombreada para fracaso], que es exactamente lo que obtuvimos cuando se usó la ecuación 7.8.

Lo que sucede es que el conocimiento adicional hace que *U* ya no sea tan relevante como espacio muestral. *Todos los enunciados de probabilidad son relativos a los espacios muestrales.* La cuestión básica, entonces, es definir adecuadamente los espacios muestrales. En el problema anterior de los esposos y esposas se hizo la pregunta: ¿Cuál es la probabilidad de que ambos, esposo y esposa sean republicanos? El espacio muestral era *U* = {*RR, RD, DR, DD*}, pero cuando se agregó el conocimiento de que uno de ellos era republicano y se elaboró la misma pregunta, se hizo al *U* original, irrelevante para el problema y un nuevo espacio muestral llamado *U'*, es requerido. Consecuentemente, la probabilidad de que ambos sean republicanos es diferente cuando tenemos más conocimiento.

Se pueden calcular otras probabilidades de forma similar. Si se deseara conocer la probabilidad de fracaso, dada una puntuación MAT menor a 65, viendo la figura 7.5, la probabilidad que se busca es la caja más grande a la derecha, etiquetada con .50. Dado que se sabe que la puntuación es <65, se utiliza este conocimiento para establecer un nuevo espacio muestral. Las dos cajas inferiores cuya área es igual a .20 + .50 = .70 representan este espacio muestral. De este modo se calcula una nueva probabilidad: .50/.70 = .71; la probabilidad de fracaso para obtener el grado si uno tiene una puntuación MAT menor a 65 es de .71.

Teorema de Bayes: revisión de las probabilidades

Ninguna discusión sobre las probabilidades condicionales podría estar completa sin mencionar brevemente el teorema de Bayes y su utilidad en la investigación de las ciencias conductuales aplicadas. Mediante la manipulación de la fórmula de la probabilidad condicional, el reverendo Thomas Bayes fue capaz de desarrollar una fórmula para calcular probabilidades condicionales especiales. Con el teorema de Bayes, se pueden actualizar o revisar probabilidades en curso, basadas en nueva información o datos. Esta información puede usarse para reestructurar la incertidumbre. Numerosos científicos e investigadores recomiendan el uso del teorema de Bayes (véase Wang, 1993).

Rara vez los datos empíricos son concluyentes. Por ejemplo, algunos estudiantes muy buenos que buscan ser admitidos en la universidad tendrán una pobre ejecución en el examen de admisión. Sin embargo, habrá algunos malos estudiantes que tendrán una buena puntuación en dicho examen. No obstante, un resultado desfavorable en un examen puede incrementar las oportunidades de rechazar a un mal estudiante, y un resultado favorable puede incrementar la probabilidad de seleccionar a un buen estudiante. De la misma forma, en la vida diaria, nosotros constantemente ajustamos viejas creencias, basados en nueva información. El teorema de Bayes, expresado cómo una fórmula numérica, muestra cómo puede hacerse esto.

$$p(C_i \mid A) = \frac{p(C_i)p(A \mid C_i)}{\sum_{j=1}^{k} p(C_j)p(A \mid C_j)}$$

Ejemplo

Basado en los datos que actualmente existen, un investigador ha determinado que el 10% de la población padece un trastorno alimenticio. Tales datos generalmente son publicados y provienen de fuentes establecidas. Lo que esto nos dice es que, sin ninguna información adicional, si 100 personas fueran elegidas al azar, 10 de ellas tendrían un trastorno alimenticio. Si T es usada para designar que una persona tiene un trastorno alimenticio, entonces $\sim T$ se usa para indicar la ausencia de dicho trastorno; por lo tanto, $p(T) = .10$ y $p(\sim T) = .90$. Éstas son llamadas "viejas" probabilidades. Algunos psicólogos se refieren a ellas como tasas base, y usar estas probabilidades por sí mismas puede llevar a resultados infructuosos.

Con la intención de mejorar la habilidad para detectar trastornos alimenticios, se desarrolló una prueba psicológica que, aunque imperfecta, da nueva información y puede ayudar a un terapeuta practicante a hacer un diagnóstico correcto. El investigador escogió individuos que se suponía presentaban este trastorno y también seleccionó un grupo de sujetos que no parecían presentarlo, y les aplicó la prueba. El número de personas con resultado positivo en la prueba, es decir, quienes en realidad tenían el trastorno, pueden representarse como una probabilidad condicional $p(+|T)$. El número de sujetos con resultado negativo en la prueba, quienes en realidad no tenían el trastorno, puede escribirse como $p(-|\sim T)$. Estas dos probabilidades condicionales son llamadas clasificaciones correctas. Empíricamente, $p(+|T) = .91$ y $p(-|\sim T) = .95$. De aquí que, con base en datos conocidos, la prueba es capaz de detectar adecuadamente al 91% de los que padecen el trastorno y al 95% de aquellos que no lo presentan. El número de personas que obtiene un resultado positivo, pero que en realidad no tiene el trastorno se designa como $p(+|T)$ y es llamado un falso positivo; el número de sujetos que recibe un resultado negativo pero que en realidad tiene el trastorno se anota como $p(-|T)$ y representa un falso negativo. Estas dos últimas probabilidades condicionales dan el nivel de imperfección de la prueba, y por lo tanto $p(+|\sim T)$ es .05 y $p(-|T) = .09$. Ahora, usando el teorema de Bayes, se pueden responder las siguientes preguntas: ¿Cuál es la probabilidad de que la persona realmente padezca el trastorno si su resultado fue negativo: $p(T|-)$? ¿Cuál es la probabilidad de que una persona padezca el trastorno y que haya resultado positivo en la prueba: $p(T|+)$? Así que usando las probabilidades condicionales, uno puede actualizar las "viejas" probabilidades. Si se utilizan estos números, la ecuación para el teorema de Bayes resulta:

$$p(T|+) = \frac{p(+|T)p(T)}{p(+|T)p(T) + p(+|\sim T)p(\sim T)} = \frac{.91(.10)}{.91(.10) + (.05)(.90)}$$
$$= \frac{.091}{.091 + .045} = \frac{.091}{.136} = 0.669 = 0.67$$

De manera similar:

$$p(T|-) = \frac{p(-|T)p(T)}{p(-|T)p(T) + p(-|\sim T)p(\sim T)}$$
$$= \frac{.09(.10)}{.09(.10) + (.95)(.90)} = \frac{.009}{.009 + .855} = \frac{.009}{.864} = 0.01$$

Con el uso de la prueba, si una persona obtiene una puntuación positiva, entonces existe un 67% de probabilidad de que padezca un trastorno alimenticio. Basándose en el teorema de Bayes, se ha ajustado la probabilidad de que la persona presente este trastorno,

de .10 a .67. Existe una probabilidad del 1% de que la persona que obtenga una puntuación negativa, tenga un trastorno alimenticio.

Doscher y Bruno (1981) utilizaron el teorema de Bayes en lugar de la fórmula habitual para hacer correcciones respecto de exámenes contestados mediante adivinación. Desarrollaron distribuciones de probabilidad de niveles de verdadero conocimiento, basados en el teorema de Bayes, de tal modo que con una calificación real de un examen, las tablas de calibración desarrolladas a partir del teorema, darían una estimación probabilística de la puntuación verdadera. Aquí, dicho teorema fue usado para ajustar las calificaciones de la prueba dadas por adivinación. Doscher y Bruno encontraron que el método de Bayes era más efectivo que la fórmula usada habitualmente para realizar la corrección. Estos mismos investigadores concluyeron que para niños urbanos el uso de una calificación sin ajustar, generalmente sobrestima el conocimiento del niño, lo cual puede resultar en que se le ubique en una situación de aprendizaje con tareas demasiado difíciles. Encontraron que con el uso de la fórmula habitual de corrección para la adivinación, el ajuste era demasiado grande y generalmente se ubicaba al niño en situaciones de aprendizaje de poco reto para él. Doscher y Bruno (1981, p. 488) dicen:

> Un procedimiento analítico basado en el teorema de Bayes permite la estimación probabilística de las calificaciones verdaderas de una prueba, usando información previa acerca de la distribución probable de calificaciones verdaderas y el patrón de adivinación específico de la población estudiada.

De manera similar al estudio de Doscher y Bruno, Jones (1991) introdujo el uso del teorema de Bayes en conjunción con las decisiones del consejero. Jones recomienda el análisis bayesiano, ya que una afirmación probabilística se hace sobre la persona que está siendo evaluada en lugar de sólo dar una puntuación y una interpretación. La investigación de Jones, basada en el teorema de Bayes, se enfocó en la selección de operadores, para emplearlos en un programa de rehabilitación para débiles visuales. Jones estableció los pasos que un consejero debía seguir en el programa, para usar el teorema de Bayes. El consejero iniciaría con un examen de los registros de la agencia y localizaría los datos psicométricos de los candidatos contratados como operadores. También determinaría cuáles serían eventualmente clasificados como exitosos y no exitosos. Jones etiquetó éstas como "creencias previas". Los registros también daban al consejero las puntuaciones de la Prueba de Destrezas Cognitivas (Cognitive Skills Test), y a partir de lo anterior, el consejero determinaría cuántos de los aspirantes exitosos tenían una puntuación de experto o superior y cuántos de los aspirantes exitosos tenían una puntuación inferior. Datos similares se obtendrían respecto de aquellos aspirantes que no fueran exitosos. Armado con esta información, el consejero podía ahora dar un estimado de probabilidad (usando el teorema de Bayes) del éxito de una persona a partir de que tuviera o no una puntuación de experto en la prueba. Jones continúa mostrando cómo integrar perfiles de personalidad en el proceso de clasificación.

El marco bayesiano abarca muchas áreas de investigación. Muchos de los métodos estadísticos más avanzados, tales como el análisis factorial confirmatorio, los modelos estructurales y el análisis discriminante, se basan en el enfoque bayesiano. El lector debería leer al menos uno de los siguientes artículos para adquirir más información. Estes (1991) proporciona algunos detalles adicionales, no mencionados aquí, sobre el uso del teorema de Bayes en casos criminales. Smith, Penrod, Otto y Park (1996) condujeron un experimento para evaluar la conducta de los jurados que aprendieron el uso del teorema de Bayes y la evidencia probabilística en casos legales de crímenes. Wang (1993) muestra la superioridad del teorema de Bayes sobre otros métodos en el pronóstico de negocios. Bierman, Bonini y Hausman (1991) dan detalles sobre el uso del teorema de Bayes en mercadotecnia e investigación de negocios.

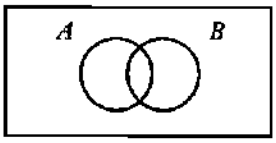
RESUMEN DEL CAPÍTULO

1. La probabilidad no es fácil de definir.
2. Hay tres definiciones amplias:
   a) La ***a priori***, implica la capacidad de la gente para dar un estimado de probabilidad en ausencia de datos empíricos.
   b) La *a posteriori* es la definición más común usada en estadística, basada en frecuencias relativas a largo plazo.
   c) El peso de la evidencia de Keynes implica dos números de probabilidad: uno es subjetivo y está basado en la cantidad de información disponible.
3. El espacio muestral es el número total de posibles resultados de un experimento. Cualquier resultado aislado es llamado punto muestral o elemento. Un evento es uno o más elementos dispuestos de tal forma que toman un significado.
4. La probabilidad de un espacio muestral es de 1.00. Los puntos muestrales y eventos son menores de 1.00. La suma de todas las probabilidades de los elementos es 1.00. A cada punto muestral se le puede asignar un peso. La suma total de los pesos usados debe ser 1.00.
5. La probabilidad tiene tres propiedades fundamentales:
   a) La medida de cualquier conjunto, como se definió antes, es mayor o igual a 0 (cero) y menor o igual a 1. En pocas palabras, las probabilidades (medidas de los conjuntos) son 0, 1, o una cifra entre ellos.
   b) La medida de un conjunto, $m(A)$ es igual a 0 si y solamente si no hay miembros en A; esto es, que A esté vacío.
   c) Suponga que A y B son conjuntos. Si A y B no están unidos —esto es $A \cap B = E$— entonces $m(A \cup B) = m(A) + m(B)$.
6. Un evento compuesto es la co-ocurrencia de dos o más eventos aislados (o compuestos).
7. La exhaustividad se refiere a la partición de un espacio muestral en subconjuntos. Cuando estos subconjuntos se combinan, cubren todo el espacio muestral.
8. Si la intersección de dos o más conjuntos resulta en un conjunto vacío, estos conjuntos son llamados mutuamente excluyentes.
9. Dos eventos se consideran independientes si la probabilidad de que ambos eventos ocurran es igual a la probabilidad de un evento multiplicado por la probabilidad del otro.
10. A veces se desea calcular la probabilidad de un evento después de recibir información adicional que puede alterar el espacio muestral. Esta probabilidad se llama probabilidad condicional.
11. El teorema de Bayes implica probabilidad condicional. Las probabilidades de Bayes son efectivas para revisar probabilidades usando información adicional o nueva.

SUGERENCIAS DE ESTUDIO

1. Suponga que usted está seleccionando una muestra de jóvenes de noveno grado para realizar una investigación. Hay 250 estudiantes de noveno grado en el sistema escolar, 130 niños y 120 niñas.
   a) ¿Cuál es la probabilidad de seleccionar a cualquier joven?
   b) ¿Cuál es la probabilidad de seleccionar a una niña? ¿y a un niño?

FIGURA 7.6



c) ¿Cuál es la probabilidad de seleccionar a un niño o una niña? ¿Cómo escribiría este problema en lenguaje de conjuntos? [Pista: ¿Es equivalente a una intersección o unión de conjuntos?]

d) Suponga que usted selecciona una muestra de 100 niños y niñas. Usted tiene 90 niños y 10 niñas. ¿Qué conclusiones podría sacar?

[Respuestas: *a)* 1/250; *b)* 120/250, 130/250; *c)* 1.]

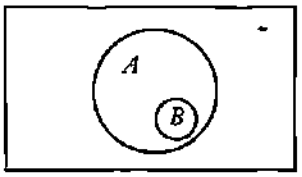
2. Lance una moneda y un dado una vez. ¿Cuál es la probabilidad de obtener cara en la moneda y 6 en el dado? Dibuje un árbol que muestre todas las probabilidades. Etiquete las ramas del árbol con los pesos o probabilidades apropiados. Ahora conteste algunas preguntas. ¿Cuál es la probabilidad de tener:
   a) cruz y un 1, un 3 o un 6?
   b) cara y un 2 o un 4?
   c) cara o cruz y un 5?
   d) cara y cruz y un 5 o un 6?

   [Respuestas: *a)* 1/4; *b)* 1/6; *c)* 1/6; *d)* 1/3.]
3. Lance una moneda y ruede un dado 72 veces. Escriba los resultados, uno al lado del otro, en una hoja cuadriculada, tal como ocurran. Verifique los resultados obtenidos contra las frecuencias esperadas teóricamente. Ahora compare sus respuestas con cada una de las respuestas de la pregunta 2. ¿Se acercan los resultados obtenidos a los resultados esperados? (Por ejemplo, suponga que usted calculó cierta probabilidad para la pregunta 2 *a)*. Ahora cuente el número de veces que la cruz se apareó con un 1, un 3, o un 6 ¿Es igual la fracción obtenida a la fracción esperada?)
4. Suponga que en la figura 7.6 hay 20 elementos en U, cuatro de los cuales están en A, seis en B y dos en $A \cap B$. Si usted selecciona aleatoriamente un elemento, ¿cuál es la probabilidad de que éste:
   a) sea de A?
   b) sea de B?
   c) sea de $A \cap B$?
   d) sea de A o de B? [Pista: recuerde la ecuación: $p(A \cup B) = p(A) + p(B) - p(A \cap B)$.]
   e) no sea ni de A ni de B?
   f) sea de B, pero no de A?

   [Respuestas: *a)* 1/5; *b)* 3/10; *c)* 1/10; *d)* 2/5; *e)* 3/5; *f)* 1/5.]
5. Observe la figura 7.6 y conteste las siguientes preguntas:
   a) Dado B, ¿cuál es la probabilidad de A?
   b) Dado A, ¿cuál es la probabilidad de B?

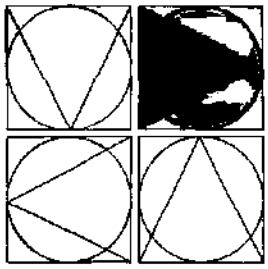
   [Respuestas: *a)* 1/3; *b)* 1/2.]
6. Considere la figura 7.7. Hay 20 elementos en U, 4 en B, y 8 en A. Si un elemento de U es seleccionado al azar, ¿cuál es la probabilidad de que el elemento sea de:
   a) A?
   b) B?
   c) $A \cap B$?
   d) $A \cup B$?
   e) U?

FIGURA 7.7



[Respuestas: *a*) 2/5; *b*) 1/5; *c*) 1/5; *d*) 2/5; *e*) 1.]

7. Con base en la figura 7.7, conteste las siguientes preguntas:
   a) Dado *A* (sabiendo que un elemento de la muestra vino de *A*), ¿cuál es la probabilidad de *B*?
   b) Dado *B*, ¿Cuál es la probabilidad de *A*?
   [Respuestas: *a*) 1/2; *b*) 1.]
8. Suponga que usted tuvo un examen de opción múltiple con dos reactivos de cuatro opciones, con las cuatro opciones de cada reactivo enumeradas *a*, *b*, *c* y *d*. Las respuestas correctas a los dos reactivos son *c* y *a*.
   a) Describa el espacio muestral. (Dibuje un diagrama de árbol; véase la figura 7.3.)
   b) ¿Cuál es la probabilidad de que cualquier examinado tenga ambos reactivos correctos por adivinación?
   c) ¿Cuál es la probabilidad de que responda correctamente al menos uno de los reactivos, por adivinación?
   [Pista: Esto puede ser un poco problemático. Dibuje un diagrama de árbol y piense en las probabilidades. Cuéntelas.]
   d) ¿Cuál es la probabilidad de contestar de forma incorrecta ambos reactivos, por adivinación?
   e) Dado que un examinado responde correctamente el primer reactivo, ¿cuál es la probabilidad de que esa persona responda correctamente el segundo reactivo, por adivinación?
   [Respuestas: *b*) 1/16; *c*) 7/16; *d*) 9/16; *e*) 1/4.]
9. La mayoría de la discusión en el texto ha sido basada en el supuesto de equiprobabilidad. Sin embargo, a menudo este supuesto no está justificado. Por ejemplo, ¿cuál es el error en el siguiente argumento? La probabilidad de que uno muera mañana es de un medio. ¿Por qué? Porque uno puede morir mañana o no morir mañana. Dado que hay dos posibilidades, cada una tiene una probabilidad de ocurrencia de un medio. ¿Cómo manejan las compañías de seguros este razonamiento? Suponga ahora, que un investigador de ciencias políticas estudia la relación entre las preferencias religiosas y políticas y asume que las probabilidades de que un católico sea republicano o demócrata fueran iguales. ¿Qué pensaría usted de sus resultados? ¿Tienen estos ejemplos implicaciones para los investigadores que conocen algo sobre los fenómenos que están estudiando? Explique.



CAPÍTULO 8

MUESTREO Y ALEATORIEDAD

Existen muchas situaciones en las que se desea saber algo acerca de las personas, de los eventos, y de las cosas. Para aprender algo acerca de la gente, por ejemplo, se selecciona gente que conocemos —o que no conocemos— y se estudia. Después del "estudio", se obtienen ciertas conclusiones, frecuentemente acerca de la gente en general. Parte de este método se basa en la sabiduría popular. Las observaciones basadas en el sentido común acerca de la gente, sus motivaciones y su comportamiento derivan, la mayoría de las veces, de observaciones y experiencias con pocas personas; se afirman hechos tales como: "la gente hoy en día no tiene valores ni sentido moral"; "los políticos son corruptos" y "los alumnos de escuelas públicas no aprenden lo suficiente". El fundamento para tales afirmaciones es: la gente saca conclusiones acerca de otras personas y acerca de su entorno, basadas principalmente en experiencias limitadas. Para llegar a tales conclusiones, las personas deben *muestrear* sus "experiencias" acerca de la gente; en realidad toman muestras relativamente pequeñas de todas las experiencias posibles. La palabra *experiencias* tiene que ser entendida en un sentido amplio, puede significar experiencia directa con otras personas, por ejemplo, una interacción de primera mano con, digamos, musulmanes o asiáticos o puede significar una experiencia indirecta, como haber escuchado hablar de musulmanes o asiáticos a los amigos o conocidos. Sin embargo, que la experiencia sea directa o indirecta, no importa mucho por ahora, ya que se asume que toda esa experiencia es directa. Un individuo afirma conocer algo acerca de los asiáticos y dice "yo sé que ellos son gregarios porque he tenido experiencia directa con muchos asiáticos" o "algunos de

mis mejores amigos son asiáticos, y yo sé que ...". El punto es que las conclusiones de esta persona están basadas en una muestra de asiáticos, o en una muestra de las conductas de los asiáticos, en o ambas. Este individuo nunca podrá "conocer" a todos los asiáticos y en su análisis debe depender de las muestras. De hecho, la mayoría del conocimiento sobre el mundo está basado en muestras, que la mayoría de las veces son inadecuadas.

Muestreo, muestreo aleatorio y representatividad

Muestrear significa tomar una porción de una población o de un universo como representativa de esa población o universo. Esta definición no dice que la muestra tomada —o extraída, como algunos investigadores dicen— sea representativa, más bien que se toma una porción de la población y ésta se *considera* representativa. Cuando un administrador escolar visita ciertos salones de clases del sistema escolar para evaluarlo, ese administrador está muestreando clases de todo el sistema escolar. Esta persona puede suponer que al visitar, digamos, de 8 a 10 salones de clases "al azar" de un total de 40, él obtendrá una clara idea de la calidad de la enseñanza que se está dando en el sistema. Otra forma podría ser visitar 2 o 3 veces la clase de un maestro para muestrear su desempeño al enseñar, con lo cual el administrador está muestreando conductas, en este caso, conductas de enseñanza, a partir del universo de todas las posibles conductas del maestro. Esta forma de muestreo es necesaria y legítima; sin embargo, pueden surgir situaciones donde el universo entero puede ser medido, entonces ¿para qué molestarse en hacer muestreos? ¿Por qué no medir cada uno de los elementos del universo? ¿Por qué no hacer un censo? Una de las principales razones es la económica. El segundo autor (HBL) trabajó en el departamento de investigación de mercado de una gran cadena de abarrotes en el sur de California, la cual constaba de 100 tiendas. La investigación de clientes y productos ocasionalmente se llevaba a cabo mediante una prueba controlada de tienda. Estos estudios fueron conducidos en una operación real diaria de una tienda de abarrotes. Quizás el interés era probar un nuevo alimento para perros, así ciertas tiendas serían elegidas para recibir el nuevo producto mientras que otro conjunto de tiendas no lo incluirían. La confidencialidad es muy importante en este tipo de estudios, ya que si un fabricante de alimento para perros de la competencia obtuviera información de que se está practicando una prueba de mercado en tal tienda, podrían contaminarse los resultados. Para llevar a cabo en una tienda pruebas controladas sobre cupones de descuento, nuevos productos o colocación en los anaqueles, podría efectuarse una investigación con dos grupos de 50 tiendas cada uno; sin embargo, el trabajo y los costos administrativos serían prohibitivos. Tendría más sentido usar muestras representativas de dicha población. Elegir 10 tiendas para cada grupo reduciría los costos de la realización del estudio. Estudios más pequeños son más manejables y controlables. Un estudio que utilice muestras puede ser realizado en un tiempo predecible. En algunas disciplinas, tales como el control de calidad y la educación (evaluación instruccional), el muestreo es esencial. En el control de calidad existe un procedimiento llamado prueba destructiva. Una manera de determinar si un producto cumple con sus especificaciones es someterlo a una prueba de rendimiento real. Cuando el producto se destruye (falla), éste puede ser evaluado. Por ejemplo, al probar neumáticos, no tendría sentido destruir cada uno de ellos para determinar si el fabricante ha cumplido con un adecuado control de calidad. De la misma manera, un maestro que quiera determinar si un niño ha aprendido el material, le aplicará un examen; sería difícil elaborar un examen que cubriera cada aspecto de lo enseñado y la retención del conocimiento del niño.

El muestreo aleatorio es el método de obtener una porción (o muestra) de una población o universo, de tal manera que cada miembro de esa población o universo tenga la

misma posibilidad de ser seleccionado. Esta definición tiene la virtud de comprenderse con facilidad; desafortunadamente, no es del todo satisfactoria debido a que es limitada. Una mejor definición es la de Kirk (1990, p. 8):

> El método de extracción de muestras a partir de una población, de manera que toda muestra posible de un tamaño particular tiene la misma posibilidad de ser seleccionada, se llama *muestreo aleatorio* y las muestras resultantes son *muestras aleatorias*.

Esta definición es general y, por lo tanto, más satisfactoria que la definición previa.

Defina un universo de estudio de todos los niños de 4° grado en cualquier sistema escolar. Suponga que son 200 niños. Ellos comprenden la población (o universo). Se selecciona un niño al azar; su posibilidad de ser seleccionado es 1/200, si el procedimiento de muestreo es aleatorio. De la misma manera se seleccionan otros niños. Suponiendo que después de seleccionar un niño, ese niño (o un símbolo asignado al niño) es regresado a la población, entonces la posibilidad de seleccionar cualquier segundo niño es también de 1/200; si no se regresa este niño a la población, entonces la posibilidad para cada uno de los niños restantes es, por supuesto, de 1/199. Esto es llamado *muestreo sin reemplazamiento*. Cuando los elementos de la muestra son regresados a la población después de haber sido elegidos, el procedimiento se llama *muestreo con reemplazamiento*.

Suponga que de la población de los 200 niños de 4° grado en cualquier sistema escolar, se obtiene una muestra aleatoria de 50 niños. Si la muestra es aleatoria, todas las muestras posibles de 50 niños tienen la misma probabilidad de ser seleccionadas —un número muy grande de muestras posibles—. Para hacer esta idea comprensible, suponga una población consistente en 4 niños *a*, *b*, *c* y *d* de donde se extrae una muestra de 2 niños al azar; entonces la lista de todas las posibilidades o *el espacio muestral* es: (*a*, *b*), (*a*, *c*), (*a*, *d*), (*b*, *c*) (*b*, *d*), (*c*, *d*). Existen 6 posibilidades. Si la muestra de 2 es seleccionada al azar, entonces su probabilidad es de 1/6, es decir, que cada uno de los pares tienen la misma probabilidad de ser elegido. Este tipo de razonamiento es necesario para resolver muchos problemas de investigación, pero generalmente tiende a limitarse a la idea más simple de muestreo asociada con la primera definición. La primera definición, entonces, es un caso particular de la segunda definición general —el caso especial en el que $n = 1$—.

Desafortunadamente, nunca se puede estar seguro de que una muestra aleatoria sea representativa de la población de la cual fue seleccionada. Recuerde que cualquier muestra particular de tamaño n tiene la misma probabilidad de ser seleccionada que cualquier otra muestra del mismo tamaño, por lo que una muestra particular puede no ser representativa en absoluto. Pero, ¿qué significa "representativa"? Por lo general, *representativo* significa que es típico de una población, es decir, que ejemplifica las características de la población. Desde el punto de vista de la investigación, *representativo* debe ser definido con mayor precisión, aunque con frecuencia es difícil ser preciso. Es necesario preguntar ¿de qué características se está hablando? Por lo tanto, en investigación, una *muestra representativa* es aquella muestra que tiene aproximadamente las mismas características de la población, relevantes a la investigación en cuestión. Si sexo y clase socioeconómica son variables (características) relevantes a la investigación, una muestra representativa tendrá aproximadamente la misma proporción de hombres y mujeres, y de individuos de clase media y clase trabajadora que la población. Cuando se selecciona una muestra al azar, se espera que sea representativa; se espera que las características relevantes de la población estén presentes en la muestra en, aproximadamente, la misma forma en que están presentes en la población, pero nunca se puede estar seguro. No hay garantía.

Como señala Stilson (1966), uno se basa en el hecho de que las características típicas de una población son aquellas más frecuentes y, por lo tanto, las que tienen mayor probabilidad de estar presentes en cualquier muestra aleatoria. Cuando el muestreo es al azar, la

variabilidad muestral es predecible. En el capítulo 7 se aprendió que, por ejemplo, si se lanzan dos dados un cierto número de veces, la probabilidad de que salga un 7 es mayor que la de que salga un 12 (véase tabla 7.1).

Una muestra extraída al azar no está sesgada, en el sentido de que ningún miembro de la población tiene más posibilidad de ser seleccionado que cualquier otro miembro; es democrática ya que todos los miembros son iguales al momento de la selección. En un ejemplo de investigación, diferente al de monedas y dados, suponga que tenemos una población de 100 niños. Los niños son diferentes en cuanto a su inteligencia, una variable relevante para el estudio. Se busca conocer la media de la puntuación de inteligencia de la población, pero por alguna razón, solamente se puede muestrear a 30 de los 100 niños. Si se muestrea aleatoriamente, hay un gran número de posibles muestras de 30 alumnos y las muestras tienen las mismas posibilidades de ser seleccionadas. Las medias de la mayoría de las muestras estarán relativamente cerca de la media de la población y algunas pocas no lo estarán. Si el muestreo fue aleatorio, la probabilidad de seleccionar una muestra con una media cercana a la media poblacional es mayor que la probabilidad de seleccionar una muestra con una media alejada de la media poblacional.

Si la muestra no es seleccionada al azar, algún factor o factores desconocidos pueden predisponer a la selección de una muestra sesgada. En este caso, quizás una muestra con una media que no esté cercana a la media poblacional. La inteligencia promedio de esta muestra será, entonces, una estimación sesgada de la media poblacional. Si se conociera a los 100 niños, inconscientemente podría tenderse a seleccionar a los niños más inteligentes: no es tanto que *nosotros lo hiciéramos* así, sino que nuestro método nos *permite* hacerlo así. Los métodos aleatorios de selección impiden que operen los propios sesgos o cualquier otro factor sistemático de selección. El procedimiento es objetivo, ya que es ajeno a las propias predilecciones y sesgos.

El lector puede estar experimentando una sensación vaga e inquietante de intranquilidad. Si no se puede estar seguro de que las muestras aleatorias sean representativas, ¿cómo confiar en los resultados de la investigación y en su aplicabilidad a las poblaciones de donde se obtuvieron las muestras? ¿Por qué no seleccionar las muestras de manera sistemática, para que sean representativas? La respuesta es compleja: primero —y de nuevo— nunca se puede estar seguro: y segundo, es más probable que las muestras aleatorias incluyan las características típicas de la población si las características son frecuentes en dicha población. En investigaciones reales se seleccionan muestras aleatorias siempre que sea posible y se espera y supone que las muestras sean representativas. Uno aprende a vivir con incertidumbre, pero procura reducirla siempre que sea posible, tal como uno lo hace en la vida diaria, pero de manera más sistemática y con suficiente conocimiento y experiencia respecto al muestreo aleatorio y a los resultados del azar. Por fortuna la falta de certeza no impide que la investigación funcione.

Aleatoriedad

El concepto de aleatoriedad está en el centro de los métodos probabilísticos modernos en las ciencias naturales y del comportamiento, pero es difícil definir *aleatorio*. El concepto del diccionario de fortuito, accidental, sin dirección o propósito, no ayuda mucho. De hecho los científicos son muy sistemáticos acerca de la aleatoriedad, ya que seleccionan con cuidado muestras aleatorias y planean procedimientos aleatorios.

Se puede asumir la postura de que nada sucede al azar, que para cualquier evento hay una causa. La única razón, de acuerdo a esta postura, para que uno use la palabra *aleatorio*, es que el ser humano no sabe lo suficiente. Para el sabio nada es aleatorio. Suponga que un

sabio tiene un periódico sabio; dicho periódico es gigantesco, y en él cada evento está cuidadosamente descrito hasta el último detalle —para mañana, para el siguiente día, y así hasta un tiempo indefinido (véase Kemeny, 1959, p. 39)—. Aquí nada es desconocido y, por supuesto, no hay aleatoriedad. Desde este punto de vista, la aleatoriedad es ignorancia.

Basándose en este argumento, la aleatoriedad se define de forma inversa: se dice que los eventos son aleatorios si no se pueden predecir sus resultados. Por ejemplo, no hay forma conocida para ganar un volado con una moneda. Debido a que no hay un sistema para jugar un juego que asegure ganarlo (o perderlo), entonces los eventos (resultados del juego) son aleatorios. Dicho de manera formal, ***aleatoriedad*** significa que no hay ley conocida capaz de ser expresada en lenguaje, que describa o explique correctamente los eventos y sus resultados. En otras palabras, cuando los eventos son aleatorios no pueden predecirse individualmente. Sin embargo, y por extraño que suene, se puede predecir con éxito en su conjunto; esto es, que se pueden predecir los resultados de un número grande de eventos. No se puede saber si el resultado de lanzar una moneda al aire será cara o cruz, pero si se lanza una moneda nivelada mil veces, se puede predecir con bastante certeza el número total de caras y cruces.

Un ejemplo de muestreo aleatorio

Para dar al lector una idea de aleatoriedad y muestras aleatorias, se utilizará una tabla de números aleatorios. Una tabla de números aleatorios contiene números generados mecánicamente, de tal manera que no tienen ningún orden perceptible o sistemático en ellos. Se dijo antes que si los eventos son aleatorios no pueden ser predichos. Pero ahora se hará una predicción de la ***naturaleza general*** de los resultados de un experimento. Mediante un cuadro de números aleatorios se seleccionan 10 muestras de 10 números cada una. Dado que los números son aleatorios, cada muestra "deberá" ser representativa del universo de números. El universo puede definirse de varias formas. Aquí se define como un conjunto completo de números en la tabla de la Rand Corporation.[1] Ahora, se extraen muestras de la tabla. Las medias de las 10 muestras serán, por supuesto, diferentes. Sin embargo, deberían fluctuar dentro de un rango relativamente estrecho, con la mayoría de ellas cerca de la media de los 100 números y de la media teórica de toda la población de números aleatorios. La cantidad de números pares en cada muestra de 10, debe ser aproximadamente igual al número de pares. Con toda seguridad habrá fluctuaciones, algunas quizás extremas, pero comparativamente la mayoría serán modestas. Las muestras se presentan en la tabla 8.1.

La medias de las muestras se encuentran debajo de cada muestra. La media de U, que es la media teórica de toda la población de los números aleatorios de Rand, {0, 1, 2, 3, 4, 5, 6, 7, 8, 9}, es 4.5. La media de los 100 números, que puede ser considerada como una muestra de U, es 4.56. Ésta es, por supuesto, muy cercana a la media de U. Puede notarse que las medias de las 10 muestras varían alrededor de 4.5, siendo 2.4 la más baja y 5.7 la más alta. Solamente dos de estas medias difieren por más de 1 del 4.5. Una prueba estadística (posteriormente se aprenderán los fundamentos de tales pruebas) demuestra que las 10 medias no difieren significativamente una de otra. (La expresión "no difieren significa-

[1] La fuente de los números aleatorios fue Rand Corporation (1955). Ésta es una tabla de números aleatorios grande y cuidadosamente construida. Estos números *no* fueron generados por computadora. Existen, sin embargo, muchas otras tablas, que son suficientemente buenas para propósitos prácticos. Muchos textos actuales de estadística tienen tablas como éstas. El apéndice C, al final de este libro, contiene 4 000 números aleatorios generados por computadora.

TABLA 8.1 *Diez muestras de números aleatorios*

	1	2	3	4	5	6	7	8	9	10	
	9	0	8	0	4	6	0	7	7	8	
	7	2	7	4	9	4	7	8	7	7	
	6	2	8	1	9	3	6	0	3	9	
	7	9	9	1	6	4	9	4	7	7	
	3	3	1	1	4	1	0	3	9	4	
	8	9	2	1	3	9	6	7	7	3	
	4	8	3	0	9	2	7	2	3	2	
	1	4	3	0	0	2	6	9	7	5	
	2	1	8	8	4	5	2	1	0	3	
	3	1	4	8	9	2	9	3	0	1	
Media	5.0	3.9	5.3	2.4	5.7	3.8	5.2	4.4	5.0	4.9	Media total = 4.56

tivamente una de otra" quiere decir que las diferencias no son mayores que las que podrían ocurrir debido al azar.) Por medio de otra prueba estadística, nueve son "buenos" estimados de la media poblacional (4.5) y uno (2.4) no lo es.

Cambiando el problema de muestreo, se puede definir un universo consistente en números pares y números impares. Suponga que en el universo entero hay un número igual de ambos. En la muestra de 100 números debería haber aproximadamente 50 números nones y 50 números pares. En realidad hay 54 números nones y 46 números pares. Una prueba estadística muestra que la desviación de 4 para los números nones y 4 para los números pares no se aparta significativamente de lo esperado por el azar.[2]

De manera similar, si se muestrean seres humanos, el número de hombres y mujeres en las muestras debe aproximarse en proporción al número de hombres y mujeres en la población (si el muestreo es aleatorio y las muestras son lo suficientemente grandes). Si medimos la inteligencia de los miembros de una muestra y la puntuación media de inteligencia de la población es 100, entonces la media de la muestra debe acercarse a 100, aunque se debe tener siempre en mente la posibilidad de seleccionar una muestra desviada, como sería una con una media de 80 o menos, o de 120 o más. Las muestras desviadas ocurren, pero es menos probable que ocurran. El razonamiento es similar al utilizado para demostraciones de lanzamientos de monedas al aire. Si lanzamos una moneda al aire tres veces, es menos probable que resulten tres caras (C) o tres cruces (X), que dos caras y una cruz o dos cruces y una cara. Esto es porque U = {*CCC*, *CCX*, *CXC*, *CXX*, *XCC*, *XCX*, *XXC*, *XXX*}. Solamente hay un punto de *CCC* y uno de *XXX*, mientras que hay tres puntos con dos C y hay tres con dos X.

Aleatorización

Suponga que un investigador desea probar la hipótesis de que el consejo psicológico puede ayudar a los estudiantes con bajo rendimiento. La prueba incluye utilizar dos grupos de estudiantes con bajo rendimiento: uno que recibirá consejo psicológico y otro que no lo

[2] La naturaleza de tales pruebas estadísticas, así como el razonamiento que las sostiene, se explicará con detalle en la parte 4. El estudiante no deberá preocuparse demasiado si no capta completamente las ideas estadísticas expresadas aquí. De hecho, uno de los propósitos de este capítulo es introducir algunos elementos básicos de tales ideas.

recibirá. Naturalmente, lo desable es tener dos grupos iguales respecto a otras variables independientes que tengan un posible efecto en el rendimiento. Una forma de hacer esto es asignar aleatoriamente a los niños a ambos grupos, por ejemplo, lanzando una moneda para cada niño. Si resulta cara, el niño es asignado a un grupo, y si resulta cruz, se le asigna al otro. Note que si hubiera tres grupos experimentales, probablemente no sería útil la moneda, sino que podría utilizarse un dado con 6 caras: si resultara un 1 o un 2, se colocaría al niño en el grupo 1, los números 3 y 4 lo pondrían en el grupo 2 y los números 5 y 6 ubicarían al niño en el grupo 3. También se podría usar una tabla de números aleatorios para signar a los niños a los grupos, de manera que si se obtiene un número non, el niño se asigna al grupo 1, y si el número es par el niño es colocado en el otro grupo. El investigador puede suponer ahora que los grupos son aproximadamente iguales respecto de todas las variables independientes posibles. Mientras más grande es el grupo, más segura es tal suposición. Así como no hay garantía de no seleccionar una muestra desviada (como se explicó antes), tampoco la hay de que los grupos sean iguales o casi iguales en todas las posibles variables independientes. No obstante, puede decirse que el investigador ha usado la aleatorización para igualar los grupos, o, dicho de otro modo, para controlar las influencias sobre la variable dependiente, que no sean las de la variable independiente manipulada. Aunque se usará el término "aleatorización", muchos investigadores prefieren usar las palabras *asignación aleatoria*; este proceso implica asignar a los participantes a las condiciones experimentales en forma aleatoria. Aunque algunos creen que la asignación al azar elimina la variación, en realidad sólo la distribuye.

Un experimento "ideal" sería aquel en que todos los factores o variables que pudieran afectar el resultado experimental fueran controlados. Si todos estos factores se conocieran, en primer lugar, y se *pudieran* hacer esfuerzos para controlarlos, en segundo lugar, entonces se tendría un experimento ideal. Sin embargo, la realidad es que no podemos conocer todas las variables pertinentes, ni podemos controlarlas (aún si las conociéramos). Sin embargo, la aleatorización puede ayudar.

La *aleatorización* es la asignación de miembros de un universo a los tratamientos experimentales, de manera que para cualquier asignación a un tratamiento, cada miembro del universo tiene la misma probabilidad de ser elegido para dicha asignación. El propósito básico de la *asignación aleatoria*, como se indicó anteriormente, es repartir sujetos (objetos, grupos) a tratamientos. Individuos con diferentes características son distribuidos de manera aproximadamente igual entre los tratamientos, de modo que las variables que puedan afectar a la variable dependiente (que no sean las variables experimentales), tengan "igual" efecto en los diferentes tratamientos. No hay garantía de que esta situación deseable se alcance, pero es más probable lograrla con la aleatorización que de otra forma. La aleatorización también tiene una razón y propósito estadísticos. Si se usa la asignación alcatoria, entonces es posible distinguir entre la varianza sistemática o experimental y la varianza del error. Las variables que pueden producir un sesgo son distribuidas a los grupos experimentales de acuerdo al azar. Las pruebas de significancia estadística (que se estudiarán más adelante) lógicamente dependen de la asignación al azar. Estas pruebas son usadas para determinar si el fenómeno observado es estadísticamente diferente al efecto del azar; sin asignación aleatoria las pruebas de significancia carecen de fundamento lógico. La idea de la aleatorización parece haber sido descubierta o inventada por Sir Ronald Fisher (véase Cowles, 1989). Fue Fisher quien virtualmente revolucionó el diseño y los métodos estadísticos y experimentales usando conceptos aleatorios como parte de su influencia. Él ha sido llamado "el padre del análisis de varianza". En cualquier caso, la aleatorización y lo que se refiere como el principio de aleatorización es uno de los mayores logros intelectuales de nuestro tiempo, y no es posible sobrestimar la importancia tanto de la idea como de las medidas prácticas que vinieron a mejorar la experimentación y la inferencia.

La aleatorización quizás puede ser aclarada en tres sentidos: estableciendo los principios de la aleatorización, describiendo cómo se usan en la práctica y demostrando cómo trabaja con objetos y números. Los tres merecen la misma importancia.

El *principio de aleatorización* puede ser enunciado como sigue: dado que en los procedimientos aleatorios cada miembro de una población tiene la misma posibilidad de ser seleccionado, si se eligen sujetos que presentan ciertas características distintivas —masculino o femenino, con alta o baja inteligencia, conservador o liberal, etcétera— probablemente a la larga serán compensados por la elección de otros miembros de la población con características, en cantidad o calidad, diferentes a las de ellos. Se puede decir que es un principio práctico que sucede en general; no se puede decir que sea una ley de la naturaleza, sino que es simplemente una afirmación de lo que ocurre con mayor frecuencia cuando se usan procedimientos aleatorios.

Se dice que los sujetos son asignados al azar a grupos experimentales y que los tratamientos experimentales son asignados al azar a grupos. Por ejemplo, en el experimento citado antes, donde se prueba la efectividad del consejo psicológico en el rendimiento escolar, los sujetos pueden ser asignados aleatoriamente a dos grupos, usando números aleatorios o lanzando una moneda al aire. Cuando los sujetos han sido asignados, los grupos pueden, a su vez, ser aleatoriamente designados como grupo experimental y grupo control usando un procedimiento similar. Se encontrarán varios ejemplos de aleatorización más adelante.

Una demostración de aleatorización senatorial

Para demostrar cómo funciona el principio de aleatorización, a continuación se describe un experimento de muestreo y diseño. Se tiene una población de 100 miembros del senado de los Estados Unidos, de los que se puede obtener una muestra. En esta población (en 1993) hay 56 demócratas y 44 republicanos. Se seleccionan dos votos importantes: el asunto 266, una enmienda para prohibir el incremento a las cuotas por pastoreo; y el asunto 290, una enmienda sobre el financiamiento para abortos. Los datos usados en este ejemplo fueron tomados del *Congressional Quarterly* de 1993. Estos votos fueron importantes porque cada uno reflejaba propuestas presidenciales. Un voto en contra en el asunto 266 y un voto a favor en el asunto 290, indicaban apoyo del presidente. Aquí se ignora la sustancia y se considera a los votos o mejor dicho, a los senadores que emitieron los votos, como poblaciones de las que tomamos la muestra.

Suponga que hacemos un experimento usando tres grupos de senadores, con 20 en cada grupo. La naturaleza del experimento no es relevante aquí. Se desea que los tres grupos de senadores sean aproximadamente iguales en todas las características posibles. Se generan números de aproximación aleatoria que van del 1 al 100, usando un programa de computadora escrito en BASIC (la referencia de los programas GWBASIC o QUICKBASIC de Microsoft se incluye en el final del capítulo). Los primeros 60 números tomados, no repetidos (muestreo sin reemplazamiento) son registrados en grupos de 20 cada uno. La afiliación política para los demócratas (d) y para los republicanos (r) se anota junto al nombre del senador. También se incluyen los votos de los senadores a las dos enmiendas: " f " para voto a favor y "c" para voto en contra. Estos datos se presentan en la tabla 8.2.

¿Qué tan "iguales" son los grupos? En la población total de 100 senadores, 56 son demócratas y 44 son republicanos, o 56% y 44%. En la muestra total de 60 hay 34 demócratas y 26 republicanos, es decir, 57% son demócratas y 43% son republicanos. Hay una diferencia de 1% para lo esperado de 56% y 44%. Las frecuencias obtenidas y las espera-

TABLA 8.2 *Voto senatorial por grupos de n = 20 en el asunto 266 del senado y el asunto 290*

#	Nombre-partido	266	290	#	Nombre-partido	266	290	#	Nombre-partido	266	290
73	hatfield-r	f	c	78	chafee-r	f	f	58	smith-r	f	c
27	coats-r	f	c	20	coverdell-r	f	c	95	byrd-d	c	c
54	kerrey-d	f	f	25	mosley-brown-d	c	f	83	matthews-d	f	c
93	murray-d	c	f	42	kerry-d	c	f	52	burns-r	f	c
6	mccain-r	f	c	68	dorgan-d	f	c	80	thurmond-r	f	c
26	simon-d	c	f	57	gregg-r	f	c	13	dodd-d	f	f
7	bumpers-d	c	f	11	campbell-d	f	f	88	hatch-r	f	c
81	daschle-d	c	f	31	dole-r	f	c	63	moynihan-d	f	f
76	specter-r	c	f	37	mitchell-d	c	f	89	leahy-d	c	f
38	cohen-r	c	f	30	grassley-r	f	c	75	wofford-d	c	f
32	kasselbaum-r	f	c	22	inouye-d	f	f	92	warner-r	f	c
44	riegel-d	c	f	99	simpson-r	f	c	91	robb-d	c	f
98	khol-d	c	f	8	pryor-d	c	?	34	mcconnell-r	f	c
77	pell-d	c	f	4	stevens-r	f	c	96	rockefeller-d	c	f
61	bingaman-d	f	c	23	craig-r	f	c	28	lugar-r	f	c
16	roth-r	c	c	12	brown-r	f	c	43	levin-d	c	f
24	kempthoner-r	f	c	10	feinstein-d	f	f	59	bradley-d	c	f
100	wallop-r	f	c	87	bennett-r	f	c	69	glenn-d	c	f
15	biden-d	c	c	19	nunn-d	c	c	9	boxer-d	c	f
14	lieberman-d	c	f	45	wellstone-d	c	f	67	conrad-d	f	c

das de los demócratas en los tres grupos (I, II y III) y la muestra total se presentan en la tabla 8.3. Las desviaciones respecto de lo esperado son pequeñas. Los tres grupos no son exactamente "iguales" en el sentido de que tengan igual número de senadores republicanos y demócratas. El primer grupo tiene 11 demócratas y 9 republicanos; el segundo grupo tiene 10 demócratas y 10 republicanos, y el tercer grupo tiene 13 demócratas y 7 republicanos. Éste no es un resultado "inusual" cuando se emplea el muestreo aleatorio. Posteriormente se verá que las discrepancias no difieren estadísticamente.

Recuerde que se está demostrando tanto el muestreo aleatorio como la aleatorización, pero especialmente la aleatorización, por lo tanto la pregunta es si la asignación aleatoria de los senadores a los tres grupos da como resultado la "igualación" de los grupos en todas sus características. Por supuesto que no se pueden probar todas las características, sino solamente las que están disponibles. En el presente caso solamente se tiene la afiliación al partido político, que se estudió anteriormente, y los votos en los dos asuntos: prohibición al incremento a las cuotas de pastoreo (asunto 266) y prohibición de fondos para ciertos tipos de aborto (asunto 290). ¿Cómo funcionó la asignación aleatoria en los dos asuntos? Los resultados se presentan en la tabla 8.4. De la votación original sobre el asunto 266, 99 de los senadores votaron a favor y 40 en contra. Estos votos totales produjeron frecuencias esperadas de votos a favor en el grupo total, de 59 ÷ 99 = .596, o 60%; por lo tanto, la expectativa es de 20 × .60 = 12 en cada grupo experimental. El voto original de los 99 senadores que votaron sobre el asunto 290 fue de 40 a favor, o 40% (40 ÷ 99 = .404). Las frecuencias esperadas para el grupo que votó a favor, entonces, son: 20 × .40 = 8. Las frecuencias obtenidas y las esperadas y las desviaciones respecto de lo esperado, para los tres grupos de 20 senadores y para la muestra total de 60 en el asunto 266 y el asunto 290 se pueden ver en la tabla 8.4.

Es notorio que todas las desviaciones de lo esperado al azar son pequeñas. Los tres grupos son aproximadamente "iguales" en el sentido de que la incidencia de los votos

TABLA 8.3 *Frecuencias obtenidas y esperadas de partido político (Demócratas) en muestras aleatorias de 20 senadores de Estados Unidos*[a]

	Grupos			*Totales*
	I	**II**	**III**	
Obtenidas	11	10	13	34
Esperadas[b]	11.2	11.2	11.2	33.6
Desviación	.2	1.2	1.8	.4

[a] Solamente se reporta mayor de las dos expectativas de la división republicano-demócrata: los demócratas (.56).
[b] Las frecuencias esperadas fueron calculadas como sigue: 20 × .56 = 11.2. De la misma forma se calculó el total: 60 × .56 = 33.6.

sobre los dos asuntos es aproximadamente la misma en cada uno de los grupos. Las desviaciones de lo esperado al azar respecto de los votos a favor (y por supuesto de los votos en contra) es pequeña. Hasta donde se puede ver, la aleatorización ha sido "exitosa". Esta demostración también puede ser interpretada como un problema de muestreo aleatorio. Se podría preguntar, por ejemplo, si las tres muestras con 20 sujetos, y la muestra total de 60 son representativas. ¿Reflejan con precisión las características de la población de 100 senadores?. ¿Reflejan las muestras la proporción de demócratas y republicanos en el senado? Las proporciones en las muestras fueron .55 y .45 (I), .50 y .50 (II), .65 y .35 (III). Las proporciones reales son .56 y .44. Aunque hay una desviación de 1%, 6% y 9% respectivamente en las muestras, estas desviaciones están dentro de lo esperado al azar. Se puede decir, por lo tanto, que las muestras son representativas en lo que respecta a la pertenencia al partido político. Un razonamiento similar se puede aplicar a las muestras y a los votos en los dos asuntos.

Ahora se puede realizar el experimento suponiendo que los tres grupos son "iguales"; por supuesto que podrían no serlo, pero las probabilidades están a favor. Y como se ha visto, el procedimiento por lo general trabaja bien. La verificación de las características de los senadores en los tres grupos mostró que los grupos son bastante "iguales" respecto a la preferencia política y a los votos a favor (y en contra) en los dos asuntos. Así se puede tener mayor confianza en que si los grupos son desiguales, las diferencias probablemente se deban a la manipulación experimental y no a las diferencias entre los grupos antes de haber iniciado la investigación.

Sin embargo, un experto como Feller (1967, p. 29), escribió:

TABLA 8.4 *Frecuencias esperadas y obtenidas de votos a favor sobre el asunto 226 y el asunto 290 en grupos aleatorios de senadores*

	Grupos						*Total*	
	I		*II*		*III*			
	266	**290**	**266**	**290**	**266**	**290**	**266**	**290**
Obtenidas	9	10	13	9	11	9	33	28
Esperadas[a]	12	8	12	8	12	8	36	24
Desviación	3	2	1	1	1	1	3	4

[a] Las frecuencias esperadas fueron calculadas para el grupo I (asunto 266), de la siguiente manera: hubo 59 votos a favor, de un total de 99 votos o 59/99 = .60; 20 × .60 = 12. Para el grupo total el cálculo es: 60 × .60 = 36.

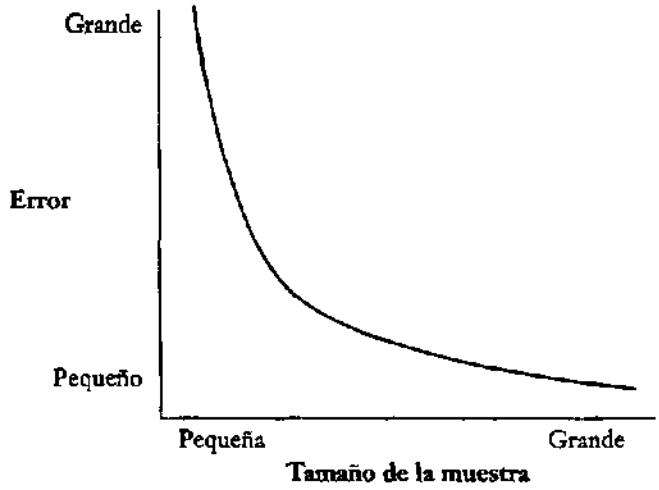
> En el muestreo de poblaciones humanas, el estadístico encuentra dificultades considerables y frecuentemente impredecibles, y la amarga experiencia ha mostrado que es difícil obtener, tan siquiera, una ordinaria imagen del azar.

Williams (1978) expone muchos ejemplos donde la "aleatorización" no funcionaba en la práctica. Uno de tales ejemplos, que influyó la vida de un gran número de hombres fue el sorteo del servicio militar en 1970. Aunque nunca fue probado del todo, parecía que los números en el sorteo no estaban aleatorizados; se metieron en una cápsula el mes y día de nacimiento, incluyendo todos los 366 días del año. Las cápsulas se colocaron en una tómbola, la cual se giraba muchas veces de modo que las cápsulas quedaran bien mezcladas. La primera cápsula seleccionada indicaba a los primeros que serían reclutados, la segunda cápsula seleccionada indicaba a los siguientes y así sucesivamente. Los resultados mostraron que las fechas para los últimos meses del año tenían una mediana menor que los primeros meses, por lo que los hombres con fecha de nacimiento hacia el final del año fueron llamados al servicio más pronto. Si la selección hubiese sido completamente al azar, las medianas para cada mes debían haber sido mucho más parecidas. El punto importante aquí es que muchos análisis estadísticos dependen de una aleatorización exitosa. Pero lograr esto en la práctica no es tarea fácil.

Tamaño de la muestra

Un regla dura pero eficaz, que se enseña a estudiantes principiantes de investigación es: utilizar una muestra tan grande como sea posible. Siempre que se calcula una media, un porcentaje o cualquier otro estadístico a partir de una muestra, se está estimando un valor poblacional. Una pregunta que debe hacerse es: ¿Qué tanto error es posible que resulte en estadísticos calculados de muestras de diferentes tamaños? La curva de la figura 8.1 expresa aproximadamente las relaciones entre el tamaño de la muestra y el error (el error es la desviación respecto a los valores poblacionales). La curva dice que a menor tamaño de la muestra, mayor será el error, y que a mayor tamaño de la muestra, menor será el error resultante.

FIGURA 8.1



Considere el siguiente ejemplo extremo. El Dr. Stanley Sue facilitó al segundo autor de este texto el puntaje de admisión de la Escala de Evaluación Global (GAS, por las siglas en inglés del Global Assessment Scale), y los días totales en terapia, de 3166 niños del condado de Los Ángeles (en las instalaciones del Centro Mental de la localidad), en los años 1983 a 1988. El doctor Sue es profesor de psicología y Director del Centro Nacional de Investigación en Salud Mental para asiático-americanos en la Universidad de California, Davis. El Dr. Sue otorgó permiso para utilizar sus datos. La información contenida en las tablas 8.5 y 8.6 se originó a partir de estos datos. El autor agradece al Dr. Stanley Sue. El GAS es un puntaje asignado por un terapeuta a cada paciente basado en su funcionamiento psicológico, social y ocupacional. La puntuación del GAS usada en este ejemplo es la que el paciente recibió al momento de su admisión o en la primera visita a la unidad.

De esta "población" se seleccionaron aleatoriamente 10 muestras de 2 niños cada una. La selección aleatoria de estas muestras y de otras se hizo usando la función "muestra" del SPSS ([paquete estadístico para ciencias sociales] [Norusis, 1992]). Las medias muestrales fueron calculadas por medio de la rutina "descriptivos" del SPSS y se pueden observar en la tabla 8.5. Las desviaciones de las medias, respecto de la media poblacional también se incluyen en ese mismo cuadro.

Las medias del GAS van de 42.5 a 60.5 y las medias de los días totales en terapia van de 14 a 303. Las dos medias totales (calculadas para las 20 puntuaciones del GAS y para las 20 puntuaciones de días totales) son 49.9 y 106.8. Estas medias calculadas a partir de muestras pequeñas varían considerablemente. Las medias de las puntuaciones de GAS y de días totales de la población (N = 3 166) fueron de 49.29 y 83.54. Las desviaciones (Des.) de las medias del GAS presentan un amplio rango: de –6.79 a 11.21. Las desviaciones de los días totales van de –69.54 a 219.46. Con muestras muy pequeñas como éstas no se puede depender de una sola media para estimar el valor de la población. Sin embargo se puede confiar más en las medias calculadas de las 20 puntuaciones, aunque ambas tienen un sesgo hacia arriba.

TABLA 8.5 *Muestras (n = 2) de las puntuaciones GAS y de días totales en terapia de 3 166 niños; media de las nuestras y desviaciones de las medias muestrales de la población (datos del Dr. Sue)*

GAS

Muestra	1	2	3	4	5	6	7	8	9	10
	61	46	65	50	51	35	45	44	43	60
	60	50	35	55	55	50	41	47	50	55
Media	60.5	48	50	52.5	53	42.5	43	45.4	46.5	57.5
Des.	11.21	–1.29	.71	3.21	3.71	–6.79	–6.29	–3.79	–2.79	8.21

Media total (20) = 49.9
Media poblacional (3 166) = 49.29

Días totales en terapia

Muestra	1	2	3	4	5	6	7	8	9	10
	92	9	172	0	3	141	28	189	28	17
	57	58	38	70	603	110	0	51	72	398
Media	74.5	33.5	105	35	303	125.5	14	120	50	207.5
Des.	–9.04	–50.04	21.46	–48.54	219.46	41.96	–69.54	36.46	–33.54	123.96

Media total (20) = 106.80
Media poblacional (3 166) = 85.54

TABLA 8.6 *Medias y desviaciones de la media poblacional de cuatro muestras de GAS y cuatro muestras de días totales (n = 20), muestra total (n = 89) y población (N = 3 166) (datos del Dr. Sue)*

	Muestras (n = 20)				Total (n = 80)		Población (N = 3 166)
GAS	49.35	48.90	49.85	50.6	49.68		49.29
Des.	.06	.39	.56	1.31		.385	
Días totales	69.40	109.95	89.55	103.45	93.08		83.54
Des.	14.14	26.41	6.01	19.91		9.540	

Cuatro muestras aleatorizadas más, de 20 puntuaciones GAS y 20 de días totales fueron tomadas de la población: las medias de estas muestras se presentan en la tabla 8.6. Las desviaciones (Des.) de cada una de las medias de las muestras de 20 puntuaciones, en relación con la media poblacional, también aparecen en la tabla, así como las medias de la muestra de 80 puntuaciones y de la población total. Las desviaciones del GAS van de .06 a 1.31, y las desviaciones de los días totales van de 6.01 a 26.41. La media de las 80 puntuaciones de GAS es de 49.68, y la media de las 3166 puntuaciones del GAS es de 49.29. Las medias comparables de los días totales son 93.08 (n = 80) y 83.54 (N = 3 166). Estas medias son mucho mejores estimados de las medias poblacionales.

Ahora se pueden sacar algunas conclusiones: primero, los estadísticos calculados a partir de muestras grandes son más precisos (si las demás características son iguales) que los calculados a partir de las muestras pequeñas. Un vistazo a las desviaciones de las tablas 8.5 y 8.6 mostrará que las medias de las muestras de 20 puntuaciones se desviaron mucho menos de la media poblacional que las medias de las muestras de dos puntuaciones. Por otro lado, las medias de la muestra de 80 sujetos se desvió muy poco de las medias poblacionales (0.39 y 9.54).

Debe ser muy claro ahora por qué el principio de la investigación y del muestreo es: use muestras grandes.[3] Las muestras grandes no se recomiendan sólo porque los números grandes sean mejores en sí mismos, sino para permitir que el principio de la aleatorización, o simplemente del azar "funcione". Con muestras pequeñas, la probabilidad de seleccionar muestras sesgadas es mayor que con muestras grandes; por ejemplo, en una muestra aleatoria de 20 senadores seleccionada hace algunos años, los primeros 10 senadores obtenidos (de 20) fueron todos ¡demócratas! Una serie de 10 demócratas sería inusual, *¡pero puede y de hecho sucede!* Digamos que se decide realizar un experimento con sólo dos grupos de 10 sujetos cada uno. Uno de los grupos tiene 10 demócratas y el otro tiene tanto demócratas como republicanos. Los resultados podrían estar seriamente sesgados, especialmente si el experimento tuviera relación con su preferencia política o sus actitudes sociales. Con grupos grandes, digamos 30 o más sujetos, hay menos riesgo. Muchos departamentos de psicología de grandes universidades tienen un requisito de investigación para los estudiantes inscritos en una clase de introducción a la psicología. Para tales situaciones, puede ser relativamente fácil obtener muestras grandes. Sin embargo, para ciertos estudios de investigación (como los de ingeniería humana o de investigación de mercados), el costo de reclutar participantes es alto. Recuerde el estudio de Williams y Adelson comentado por Simon (1987) en el capítulo 1. Así, la regla de tener muestras grandes puede no ser apropiada para todas las situaciones de investigación. En algunos estudios, 30 o más elementos, participantes o sujetos pueden ser muy pocos, especialmente en estudios que son de

[3] La situación es más compleja de lo que este simple enunciado indica. Las muestras que son demasiado grandes pueden traer otros problemas; las razones se explicarán en un capítulo posterior.

naturaleza multivariada. Comrey y Lee (1992), por ejemplo, afirman que muestras de 50 o menos sujetos poseen una inadecuada confiabilidad para los coeficientes de correlación. De aquí que puede ser más apropiado obtener una aproximación al tamaño de muestra que se necesita. La determinación estadística del tamaño muestral para los diferentes tipos de muestras se explicará en el capítulo 12.

Tipos de muestras

La discusión del muestreo ha sido hasta ahora confinada al muestreo aleatorio simple. El propósito es ayudar al estudiante a entender los principios fundamentales, por lo que se ha hecho énfasis en la idea del muestreo aleatorio simple, ya que es el fundamento de gran parte del pensamiento y de los procedimientos de la investigación moderna. El estudiante deberá comprender, sin embargo, que el muestreo aleatorizado simple no es la única clase de muestreo usado en la investigación del comportamiento; de hecho, es poco utilizado al menos para describir las características de poblaciones y las relaciones entre tales características. Sin embargo, es el modelo en el que todo muestreo científico se basa.

Otras clases de muestras pueden clasificarse de forma general en probabilísticas y no probabilísticas (y ciertas formas mixtas). Las *muestras probabilísticas* usan alguna forma de muestreo aleatorizado en una o más de sus etapas. Las *muestras no probabilísticas* no usan el muestreo aleatorizado, por lo que carecen de las virtudes que se han discutido, pero con frecuencia son necesarias e imprescindibles. Su debilidad puede, en cierta medida, ser mitigada con el uso del conocimiento, la experiencia y el cuidado al seleccionar las muestras, y replicando los estudios con diferentes muestras. Es importante que el estudiante sepa que el muestreo probabilístico no es necesariamente superior al muestreo no probabilístico en todas las situaciones y que el muestreo probabilístico tampoco garantiza muestras más representativas del universo en estudio. En el muestreo probabilístico el énfasis radica en el método y en la teoría que lo sustenta, mientras que en muestreo no probabilístico el énfasis reside en la persona que hace el muestreo y que puede acarrear consigo complicaciones enteramente nuevas e importantes. La persona que hace el muestreo debe ser conocedora de la población que se estudia, así como del fenómeno en estudio.

Una forma de muestreo no probabilístico es el *muestreo por cuotas,* donde el conocimiento de los estratos de la población —sexo, raza, región, etcétera— se utiliza para seleccionar a los miembros de la muestra que sean representativos, "típicos" y apropiados para ciertos propósitos de investigación. Un *estrato* es la partición del universo o población en dos o más grupos que no se traslapan (mutuamente excluyentes). Se toma una muestra de cada fracción. El muestreo por cuotas deriva su nombre de la práctica de asignar cuotas o proporciones de clases de personas a los entrevistadores. Este tipo de muestreo ha sido muy utilizado en encuestas de opinión pública. Para realizar este muestreo correctamente, el investigador necesita tener una lista muy completa de las características de la población, y luego debe conocer las proporciones de cada cuota. Después de esto, el siguiente paso es recolectar los datos. Dado que las proporciones pueden diferir de cuota a cuota, se les asigna un peso a los elementos de la muestra. El muestreo por cuotas es difícil de realizar porque requiere información precisa de las proporciones de cada cuota, y esta información pocas veces está disponible.

Otra forma de muestreo no probabilístico es el *muestreo propositivo,* que se caracteriza por el uso de juicios e intenciones deliberadas para obtener muestras representativas al incluir áreas o grupos que se presume son típicos en la muestra. El muestreo propositivo es usado con mucha frecuencia en la investigación de mercados. Para probar la reacción de los consumidores ante un nuevo producto, el investigador puede distribuir el nuevo producto entre personas que se ajustan al concepto que el investigador tiene del universo.

Otro ejemplo del uso del muestreo propositivo son las encuestas políticas. Con base en los resultados de votaciones pasadas y registros de partidos políticos existentes en cierta región, el investigador, propositivamente selecciona un grupo de distritos electorales. El investigador cree que esa selección comparte las características de todo el electorado. Una presentación muy interesante de cómo esta información fue usada para ayudar a elegir a un senador de Estados Unidos por el estado de California, se puede revisar en Barkan y Bruno (1972).

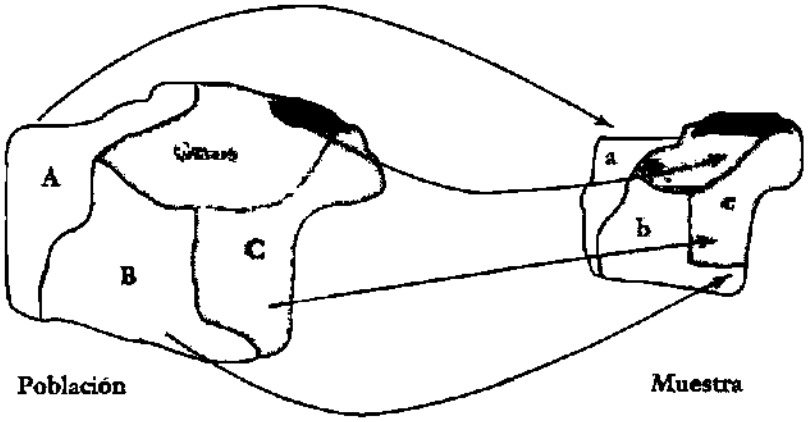
El llamado *muestreo accidental*, la forma más débil de muestreo, es quizá el más utilizado. Aquí se toman muestras disponibles a mano —estudiantes del último año de preparatoria, estudiantes universitarios de segundo año, una asociación de padres y maestros, etcétera—. Esta práctica es difícil de defender, aunque, usada con conocimiento razonable y cuidado, probablemente no merezca la mala reputación que tiene. El mejor consejo sería: evite las muestras accidentales a menos que no tenga otra opción (el muestreo aleatorio generalmente es caro y difícil de realizar); si se utilizan muestras accidentales es necesario ser extremadamente precavido en el análisis e interpretación de los datos.

El *muestreo probabilístico* incluye una variedad de formas. Cuando se explicó el muestreo aleatorio simple, se habló sobre una versión de muestreo probabilístico. Otras formas comunes de muestreo probabilístico son el muestreo estratificado, el muestreo por racimos, el muestreo por racimos de dos etapas y el muestreo sistemático. Otros métodos menos convencionales, incluyen el enfoque bayesiano o secuencial. La superioridad de un método de muestreo sobre otro se evalúa por lo regular en términos de la cantidad de variabilidad reducida en los parámetros estimados y en términos de costos. El costo es algunas veces interpretado como la cantidad de trabajo en la recolección de datos y en el análisis de datos.

En el *muestreo estratificado*, primero se divide a la población en estratos tales como hombres y mujeres, afro-americanos y mexico-americanos, etcétera. Después se seleccionan muestras aleatorias de cada estrato. Si la población consta de 52% de mujeres y 48% de hombres, una muestra estratificada de 100 participantes consistirá en 52 mujeres y 48 hombres. Las 52 mujeres se seleccionarían al azar del grupo disponible de mujeres y los 48 hombres se obtendrían aleatoriamente del grupo de hombres. Esto es llamado también distribución proporcional. Cuando este procedimiento se realiza correctamente, es superior al muestreo aleatorio simple. Comparado con el muestreo aleatorio simple, el muestreo estratificado generalmente reduce tanto la cantidad de variabilidad como el costo de recolección y análisis de datos. El muestreo estratificado saca provecho de las diferencias entre estratos. La figura 8.2 ilustra la idea básica del muestreo estratificado, el cual añade control al proceso de muestreo disminuyendo la cantidad de error de muestreo. Este diseño se recomienda cuando la población está compuesta de conjuntos de grupos desiguales. El muestreo estratificado aleatorizado ayuda a estudiar las diferencias en los estratos; permite dar especial atención a ciertos grupos que, de otra forma, podrían ser ignorados a causa de su tamaño. El muestreo aleatorizado estratificado se lleva a cabo con procedimientos de distribución proporcional (PDP). Al utilizar estos procedimientos, la división proporcional de la muestra asemeja la de la población. La mayor ventaja de usar PDP es que proveen de una muestra "auto-ponderada".

El *muestreo por racimos* —el método usado con mayor frecuencia en encuestas— es el muestreo aleatorio consecutivo de unidades o conjuntos y subconjuntos. Un racimo puede ser definido como un grupo de cosas de la misma clase; es un conjunto de elementos muestrales unidos por alguna(s) característica(s) en común. En el muestreo por racimos, el universo es fraccionado en racimos, y después los racimos son muestreados aleatoriamente. Entonces cada elemento en el racimo elegido es medido. En investigación sociológica, el investigador puede usar las manzanas de la ciudad como racimos: las manzanas de la ciu-

FIGURA 8.2



dad son elegidas aleatoriamente y los entrevistadores entonces hablan o entrevistan con todas las familias de las manzanas seleccionadas. A este tipo de muestreo algunas veces se le llama *muestreo de área*. Si un investigador utilizara el muestreo aleatorio simple o el muestreo aleatorio estratificado, necesitaría una lista completa de las familias o casas habitación para tomar su muestra. Dicha lista sería muy difícil de obtener en una gran ciudad, y aun cuando se tuviera, el costo del muestreo sería muy alto ya que implicaría medir casas habitación en una gran área de la ciudad. El muestreo por racimos es más efectivo si se usa un gran número de pequeños racimos. En investigación educacional por ejemplo, los distritos escolares de un estado o de un condado pueden usarse como racimos, tomando una muestra aleatoria de dichos distritos; cada escuela dentro del distrito escolar sería medida. Sin embargo, el distrito escolar puede conformar un racimo tan grande, que sería mejor usar las escuelas como racimos.

El *muestreo por racimos de dos etapas*, se inicia con un muestreo por racimos, como se describió antes; después, en lugar de medir cada elemento de los racimos elegidos al azar, se selecciona una muestra aleatoria de los elementos y se miden estos elementos. En el ejemplo educativo expuesto antes, se identificaría cada distrito escolar como un racimo y después se elegirían k distritos escolares al azar. De estos k distritos escolares, en lugar de medir cada escuela en los distritos elegidos (como se haría en el muestreo por racimos regular), se tomaría otra muestra aleatoria de las escuelas de cada distrito y se medirían solamente las escuelas elegidas.

Otra clase de muestreo probabilístico —si en realidad puede llamarse muestreo probabilístico— es el *muestreo sistemático*. Este método es una ligera variación del muestreo aleatorio simple. Este método supone que el universo o población consiste en elementos que están ordenados de alguna forma. Si la población consta de N elementos y se desea elegir una muestra de tamaño n, primero es necesario formar la razón N/n. Esta razón se redondea a un número entero k, el cual se usa como el intervalo del muestreo. El primer elemento muestral es elegido aleatoriamente de entre 1 y k, y los elementos subsecuentes se eligen cada k intervalos. Por ejemplo, si el elemento seleccionado aleatoriamente de los que van del 1 al 10, es 6, entonces los elementos subsecuentes son 16, 26, 36, etcétera. La representatividad de la muestra elegida de esta manera depende del ordenamiento de los N elementos de la población.

El estudiante que más adelante realizará investigación, deberá conocer mucho más sobre estos métodos, por lo que se le invita a que consulte una o más de las excelentes referencias sobre el tema, presentadas al final de este capítulo. Williams (1978) da una interesante presentación y demostración de cada método de muestreo usando datos artificiales.

Otro tema relacionado con la aleatorización y el muestreo, son las pruebas de aleatorización o permutación. Este tema se abordará de nuevo cuando se hable del análisis de datos para los diseños cuasi experimentales. Edgington (1980, 1996) fue quien propuso este método en psicología y en las ciencias del comportamiento. Él propone el uso de las pruebas de aleatorización aproximada para el análisis estadístico de los datos de muestras no aleatorias y para diseños de investigación de caso único. Ahora se explicará brevemente cómo funciona este procedimiento. Tomemos el ejemplo de Edgington de correlacionar las puntuaciones de CI de padres adoptivos y sus hijos adoptados. Si la muestra no se selecciona aleatoriamente, podría estar sesgada a favor de los padres que deseaban que se midiera su CI y el de sus hijos adoptados. Es probable que algunos padres adoptivos bajaran sus puntuaciones para hacerlas coincidir intencionalmente con el CI de sus hijos adoptados. Una forma de manejar datos no aleatorios como éste, es primero calcular la correlación entre los padres y los hijos; entonces se podrían aparear aleatoriamente las puntuaciones de los padres con las de los hijos: esto es, el papá 1 podría aparearse aleatoriamente con el hijo del padre número 10. Después de este apareo aleatorio, la correlación es calculada de nuevo. Si el investigador realiza 100 apareamientos aleatorizados y calcula la correlación cada vez, podrá entonces comparar la correlación original con los 100 apareamientos creados aleatoriamente. Si la correlación original es la mejor (la más alta), el investigador tendrá una mejor idea de que la correlación obtenida pueda ser creíble. Esta pruebas de permutación o de aleatorización han sido muy útiles para ciertas investigaciones y ciertas situaciones de análisis de datos. Se han usado para evaluar los conglomerados obtenidos en un análisis de conglomerados (véase Lee y MacQueen, 1980) y se han propuesto como una solución para el autoanálisis de eficacia de los datos que no son independientes (véase Cervone, 1987).

El azar, la aleatorización y el muestreo aleatorizado están entre las grandes ideas de la ciencia, como se indicó antes. Aunque la investigación puede, por supuesto, realizarse sin usar las ideas de la aleatorización, es difícil concebir cómo podría ser viable y tener validez, al menos en la mayoría de los aspectos de la investigación científica del comportamiento. Los conceptos modernos del diseño de investigación, del muestreo y de la inferencia, por ejemplo, son literalmente inconcebibles sin la idea del azar. Una de las paradojas más relevantes es que a través de la aleatorización o "desorden", se puede tener control sobre las complejidades con frecuencia escandalosas de los fenómenos psicológicos, sociológicos y educativos. Se impone orden al explotar las conductas conocidas de los conjuntos de los eventos azarosos. Uno queda siempre maravillado de lo que se puede llamar la belleza estructural de la probabilidad, del muestreo y de la teoría del diseño y también de su gran utilidad para resolver problemas difíciles del diseño experimental, de la planeación y del análisis e interpretación de datos.

Antes de abandonar este tema, regresemos al punto de vista sobre la aleatorización mencionado antes. Para un sabio no existe lo aleatorio. Por definición, dicho sabio "conocería" la ocurrencia de cualquier evento con completa certeza. Como Poincare (1952/1996) señala, jugar con tal sabio sería una aventura perdedora, de hecho, no sería juego. Si una moneda se lanzara 10 veces, él podría predecir caras o cruces con completa certeza y precisión. Si se lanzaran los dados, este sabio sería infalible respecto de los resultados. ¡Cada número en una tabla de números aleatorios sería predicho correctamente! Obviamente este sabio no necesitaría la investigación y la ciencia. Lo que parece afirmarse con esto es

que aleatorio es un término para la ignorancia. Si nosotros, como el sabio, conociéramos todas las causas que contribuyen a los eventos, entonces no existiría el azar. La belleza de esto, como se dijo antes, es que usamos esta "ignorancia" y la convertimos en conocimiento. Cómo se hace esto será más y más evidente conforme se avance en el estudio.

Algunos libros sobre muestreo

Babbie, E.R. (1990). *Survey research methods* (2a. ed.) Belmont, California: Wadsworth.

Babbie, E.R. (1995). *The practice of social research* (7a. ed). Belmont, California: Wadsworth.

Cowles, M. (1989). *Statistics in psychology: A historical perspective.* Hillsdale, Nueva Jersey: Erlbaum.

Deming, W.E. (1966). *Some theory of samplig.* Nueva York: Dover.

Deming, W.E. (1990). *Sampling design in business research.* Nueva York: Wiley.

Kish, L. (1953). Selection of the sample, en Festinger, L., & Katz, D. (eds.), *Research methods in the behavioral sciences.* Nueva York: Holt, Rinehart and Winston (pp. 175-239).

Kish, L. (1995). *Survey samplig.* Nueva York: Wiley.

Snedecor, G. & Cochran, W. (1989). *Statistical Methods* (8a. ed). Ames, Iowa: Iowa State University Press.

Stephan, F. & McCarthy, P. (1974). *Sampling opinions.* Westport, Connecticut: Greenwood Press.

Sudman, S. (1976). *Applied sampling.* Nueva York: Academic Press.

Warwick, D. & Lininger, D. (1975). *The sample survey: Theory and practice.* Nueva York. McGraw-Hill

Williams, B. (1978). *A sampler on sampling.* Nueva York: Wiley.

RESUMEN DEL CAPÍTULO

1. El muestreo se refiere a tomar una porción de una población o universo, que sea representativa de esa población o universo.
2. Los estudios que usan muestras son económicos, manejables y controlables.
3. Uno de los métodos más populares de muestreo es el muestreo aleatorio.
4. En el muestreo aleatorio se toma una porción (o muestra) de una población o universo de manera que cada miembro de la población o universo tiene la misma probabilidad de ser elegido.
5. El investigador define la población o universo. Una muestra es un subconjunto de la población.
6. Nunca se podrá estar totalmente seguro de que el muestreo aleatorio sea representativo de la población.
7. En el muestreo aleatorio, la probabilidad de seleccionar una muestra con una media cercana a la media poblacional es mayor que la probabilidad de seleccionar una muestra con una media alejada de la media poblacional.
8. El muestreo no aleatorio puede estar sesgado, lo que aumenta las posibilidades de que la media muestral no se acerque a la media poblacional.
9. Se dice que los eventos son aleatorios si no se pueden predecir sus resultados.
10. La asignación aleatoria es otro término para aleatorización. Aquí los participantes son asignados aleatoriamente a grupos de investigación. Se usa para controlar variables indeseables.
11. Hay dos tipos de muestras: las no probabilísticas y las probabilísticas
12. Las muestras no probabilísticas no usan la asignación aleatoria, mientras que las probabilísticas usan el muestreo aleatorio.
13. El muestreo aleatorio simple, el muestreo aleatorio estratificado, el muestreo por racimos y el muestreo sistemático, son cuatro tipos de muestreo probabilístico.

14. El muestreo por cuotas, el muestreo propositivo y el muestreo accidental son tres tipos de muestreo no probabilístico.

SUGERENCIAS DE ESTUDIO

Se recomienda una serie de experimentos con fenómenos donde interviene el azar: juegos usando monedas, dados, barajas, ruleta y tablas de números aleatorios. Estos juegos, utilizados apropiadamente, pueden ayudar a aprender mucho acerca de los conceptos fundamentales de la investigación científica moderna, la estadística, la probabilidad y por supuesto, del azar. Intente resolver los problemas que se sugieren a continuación. No se desanime si encuentra que los ejercicios de este capítulo y los posteriores son laboriosos. Es necesario y útil involucrarse directamente en la rutina de ciertos problemas. Después de trabajar con los problemas que se dan, invente los suyos. Si puede crear problemas inteligentes, de seguro ya está en el camino de entender estos conceptos.

1. De una tabla de números aleatorios, tome 50 números, del 0 al 9. (Si lo desea use los números aleatorios del apéndice C.) Lístelos en columnas de 10 números cada una.
   a) Cuente el total de números impares; cuente el total de números pares. ¿Qué esperaría obtener por el azar? Compare los totales obtenidos con los totales esperados.
   b) Cuente el total de números 0, 1, 2, 4. De forma similar cuente los números 5, 6, 7, 8 y 9. ¿Cuántos del primer grupo obtuvo? ¿Cuántos del segundo grupo? Compare lo que obtuvo con lo esperado debido al azar. ¿Se aleja mucho lo esperado de lo obtenido?
   c) Cuente los números pares y los números nones en cada grupo de 10. Cuente los dos grupos de números 0, 1, 2, 3, 4 y 5, 6, 7, 8, 9 en cada grupo de 10. ¿Difieren mucho los totales de lo esperado por efecto del azar?
   d) Sume cada columna de los 5 grupos de 10 números. Divida cada suma entre 10. (Simplemente mueva el punto decimal un lugar a la izquierda.) ¿Que esperaría obtener como la media de cada grupo, si solamente estuviera "operando" el azar? ¿Qué obtuvo? Sume las cinco adiciones y divida entre 50. La media obtenida, ¿está cercana a la esperada por azar? [Pista: Para obtener lo esperado por el azar, recuerde los límites poblacionales.]
2. Éste es un ejercicio y demostración para el salón de clases. Asigne números arbitrariamente a todos los miembros de la clase, de 1 hasta N, donde N es el total de miembros de la clase. Tome una tabla de números aleatorios e inicie en cualquier parte. Pida a un estudiante, con los ojos cubiertos, que tome un lápiz y apunte hacia la página con la tabla de números aleatorios y que marque algún punto con el lápiz. Escoja n números de dos dígitos entre 1 y N (ignore los números mayores que N y los que se repiten) bajando por las columnas (o de cualquier otra forma específica). n es el numerador de la fracción n/N, que es decidida por el tamaño de la clase. Si $N = 30$, por ejemplo, permita que $n = 10$. Repita el proceso dos veces con diferentes páginas de las tablas de números aleatorios. Ahora tendrá tres grupos iguales. Si N no es divisible entre 3, elimine una o dos personas al azar. Escriba los números aleatorios en el pizarrón para los tres grupos. Pida a cada miembro de la clase que diga su estatura en centímetros y escriba estos valores en el pizarrón, separados de los números pero en los mismos tres grupos. Sume los tres conjuntos de números en cada uno de los conjuntos en el pizarrón; sume también los números aleatorios y las estaturas. Calcule las medias de los seis conjuntos de números y calcule las medias de los conjuntos totales.

a) ¿Qué tan cercanas están las medias de cada uno de los conjuntos de números? ¿Qué tan cercanas están las medias de los grupos a la media total del grupo?
b) Cuente el números de hombres y mujeres en cada uno de los grupos. ¿Se encuentran distribuidos equitativamente ambos sexos entre los tres grupos?
c) Discuta esta demostración. ¿Cuál piensa usted que es su significado para la investigación?

3. En el capítulo 6, se sugirió que el estudiante generara 20 conjuntos de 100 números aleatorios entre 0 y 100 y que calculara las medias y las varianzas. Si usted lo hizo, use los números y estadísticas de ese ejercicio. Si no lo hizo, use los números y estadísticas del apéndice C que está al final del libro.
a) ¿Qué tan cercanas a la media poblacional están las medias de las 20 muestras? ¿Está "desviada" alguna de las medias? (Usted puede juzgar esto calculando la desviación estándar de las medias y sumando y restando dos desviaciones estándar a la media total.)
b) Con base en (a), y en su juicio, ¿son "representativas" todas las medias? ¿Qué significa "representativos"?
c) Tome las medias de los grupos tercero, quinto y noveno. Suponga que 300 sujetos han sido asignados aleatoriamente a los 3 grupos y que éstas son las puntuaciones de alguna medida de importancia en un estudio que usted desea realizar. ¿Qué cree usted que podría concluir a partir de estas tres medias?

4. La mayoría de los estudios publicados en ciencias del comportamiento y educación no usan muestras aleatorias, especialmente muestras aleatorias de grandes poblaciones. En ocasiones, sin embargo, se realizan estudios basados en muestras aleatorias. Uno de tales estudios es el realizado por Osgood, Wilson, O'Malley, Bachman y Johnston (1996). Este estudio merece una lectura cuidadosa, aunque su nivel de sofisticación metodológica incluye un gran número de detalles que van más allá del dominio que usted tiene del tema. Sin embargo, trate de no desanimarse por esta sofisticación. Saque a este documento todo el provecho que sea posible, especialmente en cuanto al muestreo de una gran población de jóvenes. Más adelante en el libro retomaremos este interesante problema. Por ahora quizás la metodología no parezca tan formidable. (Al estudiar investigación, a veces es útil leer más allá de nuestra capacidad presente, pero cuidando de no hacerlo demasiado.)

   Otro estudio de muestras aleatorias de una gran población es el realizado por Voekl (1995). En este estudio el investigador da algunos detalles acerca del uso del muestreo aleatorio estratificado en dos etapas para medir el nivel de cordialidad que el estudiante percibe de su escuela.

5. La asignación aleatoria de sujetos a grupos experimentales es mucho más común que el muestreo aleatorio de sujetos. Un ejemplo de investigación particularmente bueno (excelente, de hecho), en el que los sujetos fueron asignados al azar a dos grupos experimentales, fue realizado por Thompson (1980). De nuevo, no se deje intimidar por los detalles metodológicos de este estudio. Obtenga de él lo que pueda. Note la forma en que se clasificó a los sujetos en grupos de aptitudes y luego su asignación aleatoria a los tratamientos experimentales. También regresaremos a este estudio más adelante. Para entonces, usted será capaz de entender su propósito y diseño, y estará intrigado por el manejo experimental cuidadosamente controlado de un problema educativo muy difícil: los méritos de la llamada instrucción magistral individualizada comparados con la instrucción convencional de lectura-discusión-recitación.

6. Otro ejemplo notable de selección aleatoria es el estudio realizado por Glick, DeMorest y Hotze en 1988. Este estudio es digno de tomarse en cuenta porque

tiene lugar en un escenario real fuera de los laboratorios de la universidad. Los participantes no son por fuerza estudiantes universitarios, sino que son personas de un área pública de venta de alimentos dentro del interior de un gran centro comercial. Los participantes se seleccionaron y después se asignaron a una de seis condiciones experimentales. Este artículo es fácil de leer y el análisis estadístico no va más allá del nivel estadístico elemental.

7. Otro estudio interesante que usa otra variante del muestreo aleatorio es el de Moran y McCullers (1984). En este estudio, los investigadores seleccionaron aleatoriamente fotografías de un anuario. Después agruparon al azar dichas fotografías en 10 grupos de 16 fotos. Se les pidió a estudiantes que no conocían a los estudiantes de las fotos que evaluaran a cada persona en la foto en términos de su atractivo.

Nota especial. En algunos de los estudios sugeridos más arriba y en los del capítulo 6, se dieron instrucciones de seleccionar números de las tablas de números aleatorios o de generar grupos de números aleatorios usando una computadora. Si usted tiene una microcomputadora o tiene acceso a una, puede preferir generar números aleatorios usando la función de generador de números aleatorios de la computadora. Un libro destacado y divertido para leer y aprender cómo hacer esto es el de Walter (1999) "La Guía Secreta de las Computadoras". Walter muestra cómo escribir un sencillo programa de computadora usando lenguaje BASIC, el lenguaje común a la mayoría de las microcomputadoras. ¿Qué tan "buenos" son los números aleatorios generados?("Qué tan buenos" significa "Qué tan aleatorios".) Ya que son producidos en línea con la mejor teoría y práctica contemporánea, deben ser satisfactorios, aunque no cumplan exactamente con los requerimientos de algunos expertos. En nuestra experiencia, son bastante satisfactorios y recomendamos su uso a maestros y estudiantes. Una alternativa es el uso de números aleatorios de la Corporación Rand los cuales se reproducen parcialmente en el apéndice de este libro.

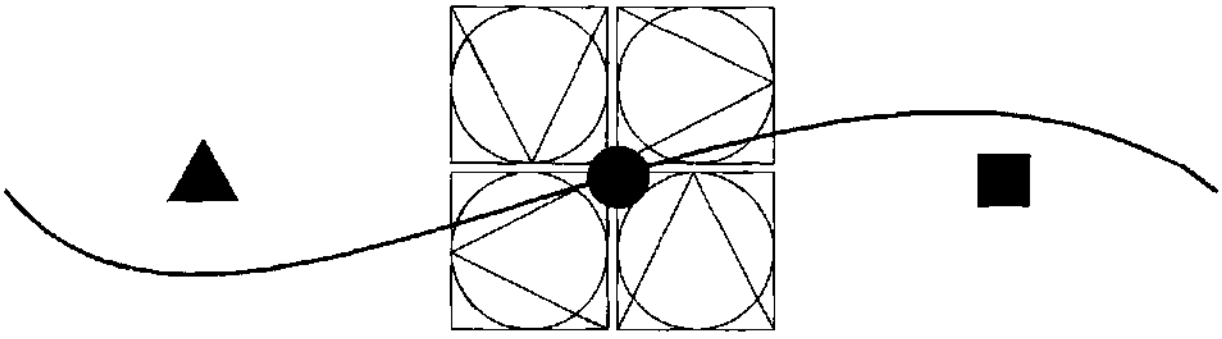
Listado de programa de cómputo para generar la tabla 8.2

```
10 DIM N(100), A$(100)
20 FOR I=1 TO 100: READ A$(I): N(I)=0
30 NEXT I
40 R=0: D=0
50 RANDOMIZE
60 I=1
70 X=RND
80 K=INT(X*100)
90 IF K=0 THEN K=100
100 IF N(K)=1 THEN 70
110 N(K)=1
120 PRINT K,A$(K)
140 Z$=RIGHT$(A$(K),1)
150 IF Z$="d" THEN D=D+1
160 IF Z$="r" THEN R=R+1
170 I=I+1
180 IF I>60 THEN 200
190 GOTO 70
200 FOR I=1 TO 100: PRINT N(I);: NEXT I
220 PRINT " ",D,R
230 DATA heflin-d,shelby-d,murkowski-r,stevens-r,deconcini-d
240 DATA mccain-r,bumpers-d,pryor-d,boxer-d,feinstein-d,campbell-d
250 DATA brown-r,dodd-d,lieberman-d,biden-d,roth-r,graham-d,mack-r
```

```
260 DATA nunn-d,coverdell-r,akaka-d,inouye-d,craig-r,kempthorne-r
270 DATA mosley-brown-d,simon-d,coats-r,lugar-r,harkin-d,grassley-r
280 DATA dole-r,kasselbaum-r,ford-d,mcconnell-r,breaux-d,johnston-d
290 DATA mitchell-d,cohen-r,mikulski-d,sarbanes-d,kennedy-d,kerry-d
300 DATA levin-d,riegel-d,wellstone-d,durenburger-r,cochran-r
310 DATA lott-r,bond-r,danforth-r,bacus-d,burns-r,exon-d,kerrey-d
320 DATA bryan-d,reid-d,gregg-r,smith-r,bradley-d,lautenberg-d
330 DATA bingaman-d,domenici-r,moynihan-d,damato-r,faircloth-r
340 DATA helms-r,conrad-d,dorgan-d,glenn-d,metzenbaum-d,boren-d
350 DATA nickles-r,hatfield-r,packwood-r,wofford-d,specter-r
360 DATA pell-d,chafee-r,hollings-d,thurmond-r,-daschle-d,pressler-r
370 DATA matthews-d,sasser-d,gramm-r,hutchinson-r,bennett-r,hatch-r
380 DATA leahy-d,jeffords-r,robb-d,warner-r,murray-d,gorton-r
390 DATA byrd-d,rockefeller-d,feingold-d,kohl-d,simpson-r,wallop-r
400 END
```

Parte Cuatro

Análisis, interpretación, estadísticas e inferencia

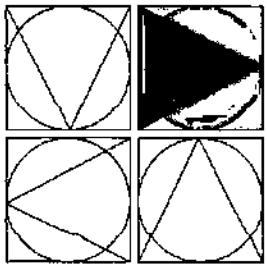


Capítulo 9
Principios del análisis e interpretación

Capítulo 10
El análisis de frecuencias

Capítulo 11
Estadística: propósito, enfoque, método

Capítulo 12
Comprobación de hipótesis y error estándar



CAPÍTULO 9

PRINCIPIOS DEL ANÁLISIS E INTERPRETACIÓN

- MEDIDAS DE FRECUENCIA Y MEDIDAS CONTINUAS
- REGLAS DE CATEGORIZACIÓN
- TIPOS DE ANÁLISIS ESTADÍSTICOS
  - **Distribuciones de frecuencia**
  - **Gráficos y elaboración de gráficos**
  - **Medidas de tendencia central y variabilidad**
  - **Medidas de relaciones**
  - **Análisis de diferencias**
  - **Análisis de varianza y métodos relacionados**
  - **Análisis de perfiles**
  - **Análisis multivariado**
- ÍNDICES
- INDICADORES SOCIALES
- LA INTERPRETACIÓN DE LOS DATOS DE INVESTIGACIÓN
  - **Adecuación de los diseños de investigación, metodología, mediciones y análisis**
  - **Resultados negativos y no concluyentes**
  - **Relaciones no hipotetizadas y hallazgos no anticipados**
  - **Prueba, probabilidad e interpretación**

El analista de la investigación descompone los datos en sus partes constituyentes para obtener respuestas a preguntas de investigación y para probar hipótesis de investigación. Sin embargo, el análisis de los datos de investigación no provee por sí mismo las respuestas a las preguntas de investigación, sino que se requiere la interpretación de los datos. Interpretar es explicar, encontrar significado. Es difícil o imposible explicar datos brutos; se debe primero analizar los datos y entonces se podrán interpretar los resultados del análisis.

Datos, como se usa en investigación del comportamiento, son los resultados de la investigación, a partir de los cuales se hacen las inferencias: generalmente son resultados numéricos, como puntuaciones de pruebas y estadísticas tales como medias, porcentajes y coeficientes de correlación. La palabra *datos* también es usada para representar los resultados de análisis matemáticos y estadísticos; pronto se estudiarán tales análisis y sus resultados. Sin embargo, los datos pueden significar algo más: puede ser información de artículos de un periódico o de una revista, material bibliográfico, diarios, etcétera, es decir, materiales verbales en general. En otras palabras, "datos" es un término general con varios significados. Piense también en los datos de investigación como los resultados de una observación y análisis sistemáticos, utilizados para hacer inferencias y sacar conclusiones. Los científicos observan, asignan símbolos y números a las observaciones y manipulan los símbolos y los números para plantearlos de forma interpretable. Posteriormente, a partir de estos datos, hacen inferencias acerca de las relaciones entre las variables de problemas de investigación.

Análisis significa la categorización, ordenamiento, manipulación y resumen de datos, para responder a las preguntas de investigación. El propósito del análisis es reducir los datos a una forma entendible e interpretable para que las relaciones de los problemas de investigación puedan ser estudiadas y probadas. Un propósito primario de la estadística, por ejemplo, es manipular y resumir los datos numéricos y comparar los resultados obtenidos con lo esperado por el azar. Un investigador hipotetiza que el estilo de liderazgo afecta la participación de los miembros del grupo de ciertas formas. El investigador planea un experimento, ejecuta el plan y recolecta datos de los sujetos. Después por medio del ordenamiento, descomposición y manipulación de los datos determina la respuesta a la pregunta ¿Cómo afectan los estilos de liderazgo a la participación de los miembros del grupo? Debe notarse que esta visión del análisis infiere que la categorización, ordenamiento y resumen de los datos debe ser planeada al inicio de la investigación. Un investigador debe trazar paradigmas de análisis o modelos aun cuando esté trabajando en el problema y en las hipótesis. Solamente de esta forma puede verse, aunque sea débilmente, si los datos y su análisis pueden de hecho contestar las preguntas de investigación.

En la *interpretación*, se toman los resultados del análisis, se hacen las inferencias pertinentes a las relaciones de investigación estudiadas y se sacan conclusiones de estas relaciones. El investigador que interpreta los resultados de investigaciones, los busca por su significado y sus implicaciones. Esto se logra de dos formas: 1) Interpretando las relaciones *dentro* del estudio de investigación y sus datos. Éste es el uso más estrecho y frecuente del término *interpretación*. Aquí, la interpretación y el análisis están estrechamente entrelazados. Casi automáticamente se interpreta cuando se hace el análisis, es decir, que cuando se calcula, digamos, un coeficiente de correlación, casi inmediatamente se infiere la existencia de una relación. 2) La búsqueda del significado más amplio de los datos de investigación. Se comparan los resultados y las inferencias realizadas a partir de los datos, con la teoría y con los resultados de otras investigaciones. Se busca el significado y las implicaciones de los resultados de la investigación dentro de los resultados del estudio y su congruencia o falta de ella, con los resultados de otros investigadores. Más importante aún, se comparan los resultados con las demandas y expectativas de la teoría.

Un ejemplo que puede ilustrar estas ideas es la investigación de cómo la forma en que se proyecta o revela cada persona hacia los demás, influye en la manera como es percibida. La teoría bajo consideración es el interpersonalismo. El interpersonalismo establece que las metas, planes y estrategias propias proveen un significado para entender a las personas y algunas de sus interacciones. Puede involucrar la construcción de modelos mentales de acción. Basados en este marco teórico general, Miller, Cooke, Tsang y Morgan (1992) predijeron que las percepciones o juicios de atribución serían determinados por la estrate-

gia de revelación que una persona decidiera adoptar. Miller *et al.*, estudiaron la diferencia entre tres tipos de revelación: negativa, positiva y jactanciosa. Se desarrollaron escenarios con diferentes métodos de revelación. Se les pidió a los participantes en el estudio que describieran su impresión de la persona en el escenario, en cinco dimensiones de atribución. Cada dimensión fue correlacionada con el tipo de escenario. La correlación (calculada en la forma de η^2[eta^2]) fue muy alta. Éste es el análisis. Los datos se delinearon en una serie de dos conjuntos de medidas, que después fueron comparadas a través de un procedimiento estadístico.[1]

El resultado del análisis —un coeficiente de correlación— debe ser ahora interpretado. ¿Cuál es su significado?, específicamente, ¿cuál es su significado dentro del estudio? ¿Cuál es su significado más amplio a la luz de los hallazgos e interpretaciones de investigaciones relacionadas? Y, ¿qué significado tiene, al confirmar o no confirmar las predicciones teóricas? Si la predicción "interna" se sostiene, entonces se relacionan los hallazgos con los de otras investigaciones, que pueden ser o no consistentes con el hallazgo presente.

La correlación fue alta, por lo que los datos de la correlación son consistentes con las expectativas teóricas. La teoría del interpersonalismo establece que las diferentes estrategias de revelación influyen en la percepción. El fanfarronear es una estrategia de revelación, por lo que ésta debe influir en la percepción. La inferencia específica es que las cosas que uno dice acerca de sí mismo influyen en la opinión que los demás tengan de esa persona. En ciertas situaciones el fanfarronear sobre sí mismo sirve para un propósito útil, la gente lo verá como confiable y exitoso. Mientras que las revelaciones negativas tenderán a hacer que la gente piense que la persona es socialmente sensible, pero no exitosa. Se miden al menos dos variables y se correlacionan. A partir del coeficiente de correlación se da un salto inferencial hacia la hipótesis. Dado que es alta (como se predijo) la hipótesis se apoya. Entonces, se intenta relacionar el hallazgo con otras investigaciones y teorías.

Medidas de frecuencia y medidas continuas

Los datos cuantitativos se presentan en dos formas generales: como medidas de frecuencia y como medidas continuas. Obviamente, las medidas continuas están asociadas con variables continuas (véase la explicación sobre variables continuas y variables categóricas en el capítulo 3). Aunque ambas clases de variables y medidas pueden ser integradas bajo el mismo marco de medición o referencia, en la práctica es necesario distinguirlas.

Frecuencias son los números de objetos en un conjunto o subconjunto. Suponga que U es el conjunto universal con N objetos. Por lo tanto N es el *número* de objetos de U. Permita que U sea fraccionada en $A^1, A^2, \ldots, A_k$. Permita que $n_1, n_2, \ldots, n_k$ sea el número de objetos en $A_1, A_2, \ldots, A_k$. Entonces $n_1, n_2, \ldots, n_k$ son llamadas frecuencias.

Resulta útil ver esto como una función. Suponga que X es cualquier conjunto de objetos con miembros $\{x_1, x_2, \ldots, x_k\}$. Se desea medir un atributo de los miembros del conjunto, el cual será llamado M. Permita que $Y = \{0, 1\}$ y que la medición sea descrita como una función: $f = \{(x, y)$; donde x es un miembro del conjunto X, y y es 1 o 0, dependiendo de que x posea o no a $M\}$. Esto se lee: f, una función o regla de correspondencia es igual al conjunto de pares ordenados (x, y), donde x es un miembro de X, y es 1 o 0, y así

[1] En el estudio de Miller, Cooke, Tsang y Morgan (1992) se empleó más que un análisis de correlación. También realizaron tanto el análisis univariado como el multivariado de varianza. η^2 mide la relación entre la variable independiente: revelación; y la(s) variable(s) dependiente(s): atribución.

sucesivamente. Si x posee M (determinado de alguna manera empírica), se le asigna un 1. Si x no posee M, se le asigna un 0. Para encontrar la frecuencia de objetos con la característica M, se cuenta el número de objetos a los que se les haya asignado el número 1.

Con medidas continuas, la idea básica es la misma. Sólo la regla de correspondencia, f, y los números asignados a los objetos cambian. La regla de correspondencia es más elaborada y los numerales son generalmente 0, 1, 2,... y fracciones de estos numerales. En otras palabras, se escribe una ecuación de medición:

$$f = \{(x, y); x \text{ es un objeto, y } y \text{ es cualquier numeral}\}$$

que es una forma generalizada de la función. (Esta ecuación y las ideas que la sustentan se explicarán en detalle en el capítulo 25.) Esta digresión es importante porque ayuda a ver las similitudes básicas del análisis de frecuencias y del análisis de medidas continuas.

Reglas de categorización

El primer paso de cualquier análisis es la categorización. Se dijo anteriormente (capítulo 4) que la partición es el fundamento del análisis. Ahora se explicará por qué. *Categorización* es un sinónimo de partición —es decir, una *categoría* es una partición o una subpartición—. Si un conjunto de objetos es categorizado de alguna forma, es fraccionado de acuerdo con una regla. La regla dice, en efecto, cómo asignar los objetos del conjunto a las particiones o subparticiones. Si esto es así, entonces las reglas de partición que se estudiaron con anterioridad se aplican a los problemas de categorización. Solamente es necesario explicar las reglas, relacionarlas con los propósitos básicos del análisis y ponerlas a trabajar en situaciones analíticas prácticas.

A continuación se describen las cinco reglas de categorización; la (2) y la (3) son las reglas de exhaustividad y de separación discutidas en el capítulo 4. Las otras (4 y 5), en realidad se pueden deducir de las reglas fundamentales (2) y (3). Por razones prácticas, se listan como reglas separadas.

1. Las categorías se establecen de acuerdo con el problema de investigación y sus propósitos.
2. Las categorías son exhaustivas.
3. Las categorías son mutuamente excluyentes e independientes.
4. Cada categoría (variable) se deriva de un principio de clasificación.
5. Cualquier esquema de categorización deberá de estar en un nivel de discurso.

La regla 1 es la más importante. Si las categorizaciones no se establecen de acuerdo a las demandas del problema de investigación, entonces no puede haber respuestas adecuadas a las preguntas de investigación. Hay una pregunta que constantemente debe hacerse: ¿Se ajusta el paradigma del análisis al problema de investigación? Suponga que la pregunta de investigación dice: ¿Cuál es la influencia de la televisión en las habilidades para procesar la comunicación no verbal de los niños? Se ha dicho que demasiada televisión es mala para los niños. ¿Es esto verdad? Cualesquiera que sean los datos obtenidos y el análisis realizado, éstos deben soportar el problema de investigación, que en este caso es la relación entre cantidad de televisión y comprensión de la comunicación no verbal.

La clase más simple de análisis es el análisis de frecuencia. Feldman, Coats y Spielman (1996), en su estudio sobre la cantidad de televisión que se ve y la comprensión de la comunicación no verbal, seleccionaron una muestra de niños y determinaron la frecuencia con que ellos veían la TV. Después, midieron la comprensión que cada niño tenía del

uso estratégico de las manifestaciones emocionales no verbales del personaje principal de un programa de TV. Feldman, Coats y Spielman dividieron a los niños en tres grupos, de acuerdo a la frecuencia de exposición a la TV, en: ligera, moderada y alta. Después contaron cuántos de estos niños eran capaces de ofrecer una respuesta simple o compleja acerca de la presentación visual de las emociones del personaje principal en el programa de TV que veían. El paradigma para el análisis de frecuencia era el siguiente:

Nivel de exposición a la televisión

Regla de Categorización Demostrada	Ligera	Moderada	Alta
Simple Compleja		Frecuencia	

Dado que Feldman *et al.* midieron la cantidad de exposición a la TV en forma continua, ellos podrían haber usado este paradigma:

Regla de Categorización Demostrada

Simple	Compleja
Cantidad de exposición a la TV	

Es obvio que los dos paradigmas soportan directamente el problema: en ambos es posible probar la relación entre la comprensión y la exposición a la televisión, si bien es cierto que de diferentes formas. Los autores eligieron el primer método —y encontraron que aquellos niños que veían menos televisión mostraban un mayor nivel de comprensión—. En el grupo de exposición ligera, el 50% de los niños dieron explicaciones complejas y heterogéneas. En el grupo de alta exposición a la TV, 0% mostró un alto nivel de comprensión. El segundo paradigma indudablemente llegaría a la misma conclusión. El punto es que un paradigma analítico es, en efecto, otra forma de formular un problema, una hipótesis y una relación. El hecho de que un paradigma utilice frecuencias mientras que el otro use medidas continuas no afecta de ninguna forma la relación probada. En otras palabras, ambas formas de análisis son lógicamente similares: ambas prueban la misma proposición pero pueden diferir en los datos usados, en pruebas estadísticas, y en sensibilidad y poder.

Hay varias cosas que un investigador podría hacer, que serían irrelevantes para el problema. Si una, dos, o tres variables son incluidas en el estudio sin una razón teórica o práctica para hacerlo, entonces el paradigma analítico sería, al menos parcialmente, irrelevante al problema. Suponga que un investigador estudia la hipótesis de que la educación religiosa aumenta el carácter moral de los niños, y para ello recolecta datos con una prueba de rendimiento en niños de escuelas públicas y de escuelas religiosas. Esto probablemente no tenga relevancia para el problema. El investigador está interesado en las diferencias morales, no en las diferencias de logro, entre los dos tipos de escuelas y entre la instrucción religiosa y la instrucción no religiosa. Podrían incluirse otras variables que tuvieran poco o nada que ver con el problema, como por ejemplo, las diferencias en la experiencia y entrenamiento del profesor o razones alumno-maestro. Si, por otro lado, el investigador piensa que ciertas variables tales como sexo, religión familiar, y quizás variables de personalidad, puedan interactuar con la instrucción religiosa para producir diferencias, enton-

ces sería justificable incorporar tales variables dentro del problema de investigación y, consecuentemente, en el paradigma analítico.[2]

La Regla 2, sobre la exhaustividad, dice que todos los sujetos u objetos de *U*, deben ser usados. Todos los individuos en el universo deben tener la posibilidad de ser asignados a las casillas del paradigma analítico. Con el ejemplo anteriormente considerado, cada niño acude a una escuela religiosa o a una escuela pública. Si, de algún modo, la muestra hubiera incluido niños que asistieran a escuelas privadas, entonces la regla habría sido violada porque habría un número de niños que no se ajustarían al paradigma del problema. (¿Cómo sería un paradigma de análisis de frecuencia? Considere a la honestidad como variable dependiente.) Sin embargo, si el problema de investigación hubiera considerado a los alumnos de escuelas privadas, entonces el paradigma tendría que cambiarse, añadiendo la categoría "privada" a las rúbricas "religiosa" y "pública".

El criterio de exhaustividad no siempre es fácil de satisfacer. Con algunas variables categóricas, no hay problema. Si el género es una de las variables, cualquier individuo ha de ser masculino o femenino. Suponga, sin embargo, que una variable bajo estudio fuera la preferencia religiosa y que se incluyera en el paradigma: Protestante-Católico-Musulmán. Ahora suponga que algunos sujetos fueran ateos o budistas. Claramente se puede observar que el esquema de categorización viola la regla de exhaustividad: algunos sujetos no tendrían casillas a donde ser asignados. Dependiendo del número de casos y del problema de investigación se podría agregar la categoría "Otras", en la cual se incluirán a los sujetos que no son ni protestantes, ni católicos, ni musulmanes. Otra solución, especialmente cuando el número de sujetos en "Otras" es pequeño, es eliminar a estos sujetos del estudio. Otra forma de solucionarlo sería ubicar a estos sujetos, si fuera posible, bajo una categoría ya existente. Algunos ejemplos de otras variables donde se encuentran estos problemas son: la preferencia política, la clase social y los tipos de educación.

La Regla 3 frecuentemente causa preocupación a los investigadores. Esta regla demanda que las categorías sean mutuamente excluyentes, lo que significa, como se aprendió antes, que cada objeto de *U*, cada sujeto de la investigación (es decir, la medición dada a cada sujeto), debe ser asignado a una casilla y solamente a una casilla de un paradigma analítico. Ésta es una función de la definición operacional. Las definiciones de las variables deben de ser claras y sin ambigüedades, de tal manera que sea poco probable que cualquier sujeto se asigne a más de una casilla. Si la preferencia religiosa es la variable que está siendo definida, entonces la definición de los subconjuntos Protestante, Católico y Musulmán debe quedar clara y sin ambigüedades. La definición podría ser: "miembro registrado en una iglesia" o también "nacido en la iglesia" y puede ser tan simple como la identificación que el propio sujeto haga de sí mismo como un protestante, un católico, o un musulmán. Cualquiera que sea la definición, debe permitirle al investigador asignar a cualquier sujeto a ***una y solamente a una*** de las casillas.

La parte de independencia de la regla 3 es a menudo difícil de satisfacer, especialmente con medidas continuas —y algunas veces con frecuencias—. ***Independencia*** significa que la asignación de un objeto a una casilla no afecte, de ninguna forma, la asignación de cualquier otro objeto a esa casilla o a cualquier otra casilla. La asignación aleatoria de un universo infinito o lo suficientemente grande, por supuesto, satisface la regla. Sin la asignación aleatoria, puede haber problemas. Cuando se asignan objetos a las casillas con base

[2] En el capítulo 6, se hicieron algunas consideraciones elementales del análisis de frecuencia con más de una variable independiente. En capítulos posteriores se hará una consideración más detallada tanto del análisis de medidas de frecuencia como del análisis de medidas continuas, con muchas variables independientes. El lector no deberá preocuparse si no logra pleno entendimiento y comprensión de los ejemplos dados anteriormente. Más adelante resultarán más claros.

en que el objeto posea ciertas características, la asignación de un objeto, ahora, puede afectar posteriormente la asignación de otro objeto.

La Regla 4, que dice que cada categoría (variable) se deriva de un principio de clasificación, algunas veces es violada por el neófito. Si se tiene una firme comprensión de la partición, este error puede evitarse fácilmente. La regla implica que, al establecer un diseño analítico, cada variable ha de ser tratada separadamente debido a que cada variable representa una dimensión separada. No se ponen dos o más variables en una categoría o en una dimensión. Si se estuvieran estudiando, por ejemplo, las relaciones entre clase social, sexo y adicción a las drogas, no se pondría a la clase social y al sexo en la misma dimensión.

Para ilustrar esto conviene citar un estudio de Glick, DeMorest y Hotze (1988). Estos investigadores estudiaron las relaciones entre la pertenencia grupal, el espacio personal y la respuesta a una solicitud de ayuda. En este estudio, cómplices de los investigadores buscaron ayuda de una persona con características físicas similares o diferentes (pertenencia a un grupo). Adicionalmente, al pedir ayuda, ellos se encontraban a poca, mediana o larga distancia (espacio personal) del sujeto. La persona a quien se acercaban respondía o no al requerimiento de ayuda. En este estudio, un error en la regla 4 podría verse de la siguiente forma:

	Dentro del grupo	Fuera del grupo	Cerca	Medio	Lejos
Accedió	Frecuencias				
Rehusó					

Claramente este paradigma viola la regla: tiene solamente una categoría derivada de dos variables. Cada variable debe tener su propia categoría. Un paradigma correcto debería verse así:

	Dentro del grupo			*Fuera del grupo*		
	Cerca	Medio	Lejos	Cerca	Medio	Lejos
Accedió	Frecuencias					
Rehusó						

La Regla 5 es la más difícil de explicar porque el término "nivel del discurso" es difícil de definir. Ya fue definido en un capítulo anterior como un conjunto que contiene todos los objetos que entran en una discusión. Si se usa la expresión "universo del discurso", se liga la idea a ideas establecidas. Cuando se habla acerca de U_1, no se trae a colación a U_2 sin una buena razón y sin haber aclarado que se está haciendo esto. Para una discusión sobre los niveles del discurso y su relevancia se puede revisar Kerlinger (1969, pp. 1127-1144, especialmente p. 1131).

El análisis de investigación generalmente mide la variable dependiente: por ejemplo, considere el problema de la pertenencia a un grupo, el espacio personal y la respuesta a una solicitud de ayuda. La pertenencia a un grupo y el espacio personal son las variables independientes; responder a una solicitud de ayuda es la variable dependiente. Los objetos de análisis son las medidas de la respuesta a la solicitud de ayuda. Las variables independientes y sus categorías son realmente usadas para estructurar el análisis de la variable dependiente. El universo del discurso, U, es el conjunto de medidas de la variable dependiente. Las variables independientes pueden percibirse como los principios de partición

usados para fraccionar las medidas de la variable dependiente. Si súbitamente cambiamos a otro tipo de medida de la variable dependiente, entonces habremos cambiado los niveles o universos del discurso.

Tipos de análisis estadísticos

Hay muchos tipos de análisis estadísticos y de presentación que no pueden exponerse en detalle en este libro. Explicaciones posteriores de ciertas formas más avanzadas de análisis estadísticos, tienen como propósito la comprensión básica de la estadística y la inferencia estadística, y la relación de la estadística y la inferencia estadística con la investigación: Aquí, las formas más sofisticadas de análisis estadísticos se exponen brevemente para dar al lector un panorama del tema; sin embargo, se explican solamente en su relación con la investigación. Se asume que el lector ya ha estudiado la estadística descriptiva más simple. Aquellos que no lo han hecho, podrán encontrar buenas explicaciones en libros de texto elementales (Comrey y Lee, 1995; Kirk, 1990; Howell, 1997; Hays, 1994).

Distribuciones de frecuencia

Aunque las distribuciones de frecuencia se usan principalmente para propósitos descriptivos, también pueden ser usadas para otros propósitos de investigación. Por ejemplo, se puede probar si dos o más distribuciones son lo suficientemente similares para garantizar su unión. Suponga que se estudia el aprendizaje verbal de niños y niñas de 6° grado. Después de obtener un gran número de puntuaciones de aprendizaje verbal, pueden compararse y probar las diferencias entre las distribuciones de niños y niñas. Si la prueba muestra que las distribuciones son iguales —y otros criterios se satisfacen— entonces quizás puedan ser combinadas para otros análisis.

Las distribuciones observadas también pueden ser comparadas con distribuciones teóricas. La comparación más conocida de este caso es la llamada *distribución normal*. Puede ser importante saber que las distribuciones obtenidas son normales en forma, o si no son normales, se desvían de la normalidad en ciertas formas específicas. Dicho análisis puede ser útil en otros trabajos teóricos y aplicados, así como en la investigación. En un estudio teórico de habilidades es importante saber si estas habilidades están, de hecho, distribuidas normalmente. Dado que se ha encontrado que muchas de las características humanas se distribuyen normalmente (ver Anastasi, 1958),[3] los investigadores pueden hacer preguntas importantes acerca de características "nuevas" que están siendo investigadas.

La investigación educativa aplicada puede valerse del cuidadoso estudio de la distribución de la inteligencia, las aptitudes y las puntuaciones de aprovechamiento. ¿Es concebible que un programa de aprendizaje innovador pueda cambiar las distribuciones de las puntuaciones de aprovechamiento, digamos, de los niños de 3° y 4° grado? ¿Podría ser que los programas de educación masiva temprana pudieran cambiar los perfiles de las distribuciones, así como los niveles generales de las puntuaciones?

Allport (1947), en su estudio sobre el conformismo social, mostró que aun un fenómeno complejo de comportamiento como el conformismo, podría estudiarse provechosa-

[3] El estudiante de investigación en educación, psicología y sociología debe estudiar la sobresaliente contribución de Anastasi a la comprensión de las diferencias individuales. Su libro también contiene muchos ejemplos de distribuciones de datos empíricos.

mente usando el análisis de distribución. Allport fue capaz de demostrar que muchas de las conductas sociales —parar frente a una luz roja, infracciones de estacionamiento, observancia religiosa, etcétera— se distribuían en forma de una curva *J*, donde la mayoría de las personas son conformistas, pero con un número pequeño predecible de inconformistas, en diferentes grados. Coren, Ward y Enns (1994) presentan numerosas distribuciones con diferentes formas para ciertas percepciones humanas de estímulos físicos basados en la Ley Psicofísica de Steven.

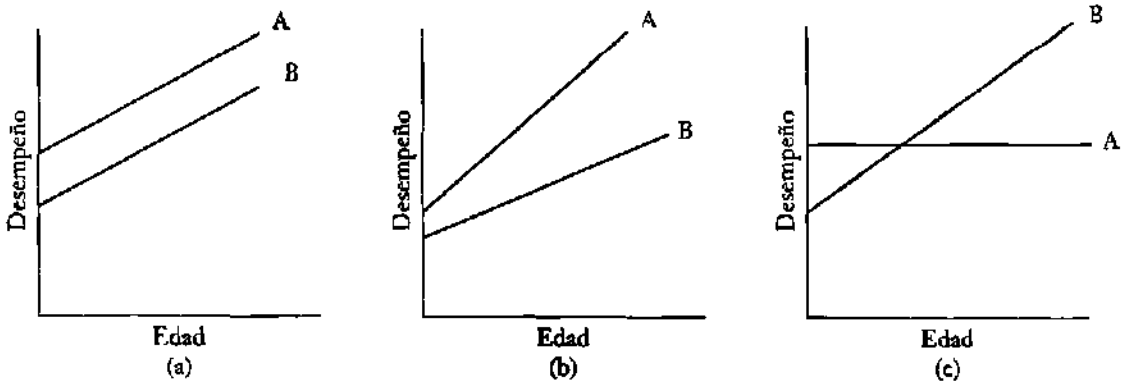
Las distribuciones han sido, probablemente, muy poco usadas en las ciencias del comportamiento y en la educación: el estudio de las relaciones y las pruebas de hipótesis son casi automáticamente asociadas con correlaciones y comparaciones de promedios. El uso de distribuciones es considerado con menos frecuencia. Algunos problemas, sin embargo, pueden ser mejor resueltos usando los análisis de distribución. Estudios de patología y de otras condiciones inusuales son quizás mejor abordados a través de la combinación de los análisis de distribución y los conceptos probabilísticos.

Gráficos y elaboración de gráficos

Una de las más poderosas herramientas del análisis es el gráfico. Un *gráfico* es una representación bidimensional de una relación o relaciones. Exhibe gráficamente conjuntos de pares ordenados en una forma que ningún otro método puede hacerlo. Si existe una relación en un conjunto de datos, un gráfico no sólo la mostrará claramente, sino que también mostrará su naturaleza: positiva, negativa, lineal, cuadrática, etcétera. Aunque los gráficos han sido usados con frecuencia en las ciencias del comportamiento, al igual que las distribuciones, al parecer no han sido lo suficientemente utilizados. Para estar seguros, hay formas objetivas de resumir y probar relaciones, tales como coeficientes de correlación, comparación de medias y otros métodos estadísticos, sin embargo, ninguno de éstos describe tan vívidamente una relación como un gráfico.

Revisando los gráficos del capítulo 5 (figuras 5.1, 5.4, 5.5 y 5.6), se puede notar cómo transmiten la naturaleza de las relaciones. Posteriormente se usarán gráficos de manera más interesante para mostrar la naturaleza de relaciones más complejas entre variables. Para dar al estudiante sólo una muestra de la riqueza de tal análisis, anticiparemos una discusión posterior; de hecho, intentaremos enseñar una idea compleja usando gráficos.

FIGURA 9.1



Los tres gráficos de la figura 9.1 muestran tres relaciones hipotéticas entre la edad, como variable independiente y el desempeño verbal (etiquetado "desempeño") como variable dependiente, de niños de clase media (A) y niños de clase trabajadora (B). Podría llamarse a estos gráficos, gráficos de desarrollo. El eje horizontal es la abscisa y se usa para indicar la variable independiente o *X*. El eje vertical es la ordenada y es usado para indicar la variable dependiente o *Y*. El gráfico (a) muestra la misma relación positiva entre edad y desempeño tanto en la muestra A, como en la muestra B. También muestra que los niños de la muestra A superan a los niños de la muestra B. El gráfico (b) muestra que ambas relaciones son positivas, pero que conforme pasa el tiempo, el desempeño de los niños de la muestra A se incrementó más que el desempeño de los niños de la muestra B. El gráfico (c) es más complejo. Muestra que los niños de la muestra A superaron a los niños de la muestra B en una etapa temprana y que se mantuvieron así hasta una edad mayor, pero los niños de la muestra B, que iniciaron más bajos, avanzaron y continuaron su avance en el tiempo hasta que superaron a los niños de la muestra A. Este tipo de relación es poco probable en el desempeño verbal, pero puede ocurrir con otras variables.

El fenómeno mostrado en los gráficos (b) y (c) es conocido como *interacción*. Brevemente, implica que dos o más variables interactúan en su "efecto" sobre una variable dependiente. En este caso, la edad y el estatus del grupo interactúan en su relación con el desempeño verbal. Expresado de otra forma, la interacción significa que la relación de una variable independiente con una variable dependiente difiere en grupos distintos, como en este caso, o a diferentes niveles de otra variable independiente. El estudio de Behling y Williams (1991) arrojó resultados que podrían ser graficados como en el gráfico (b). En este estudio los investigadores examinaron la percepción del estudiante y del maestro de la inteligencia a partir de diferentes estilos de vestuario de hombres y mujeres. Un gráfico de un estilo de vestuario está en la figura 9.2. Un gráfico similar al gráfico (c) puede construirse a partir de los datos proporcionados por Little, Sterling y Tingstrom (1996). Su estudio involucra la percepción que tenían estudiantes del norte y del sur de Estados Unidos de una persona objetivo descrita, ya sea como norteña o como sureña. Así, los estudiantes del sur dieron puntuaciones similares en el diferencial semántico, tanto a las personas del norte como a las del sur. Sin embargo, los estudiantes del norte dieron a las personas del norte puntuaciones mucho más altas que a aquellas del sur. Esto se ilustra en la figura 9.3. La noción de un efecto de interacción se explicará en detalle y con precisión cuando se estudie el análisis de varianza y el análisis de regresión múltiple.

FIGURA 9.2

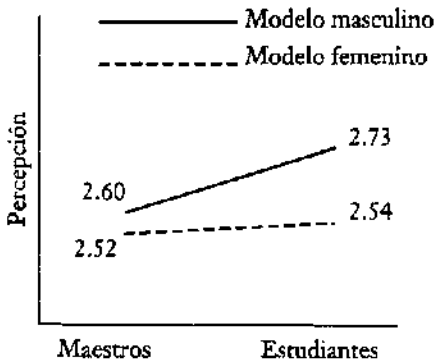
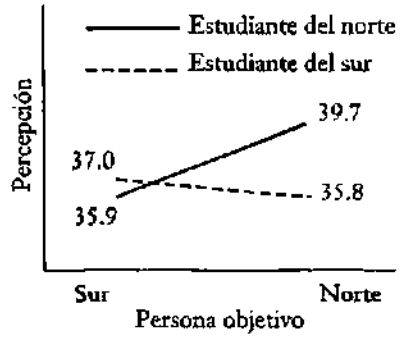


FIGURA 9.3



Mientras que las medias representan una de las mejores formas para reportar datos complejos, confiar completamente en ellos puede ser desafortunado. La mayoría de los casos de diferencias significativas de medias entre grupos, también se acompañan por un considerable traslape de las distribuciones. Anastasi da ejemplos claros (1958) y señala la necesidad de prestar atención a los traslapes y da ejemplos y gráficos de las diferencias de la distribución de sexo, entre otras. En pocas palabras, se les aconseja a los estudiantes de investigación que adquieran el hábito, desde el inicio de su estudio, de prestar atención y comprender las distribuciones de variables y de graficar relaciones de variables.

Medidas de tendencia central y variabilidad

Prácticamente no hay duda de que las medidas de tendencia central y variabilidad son las herramientas más importantes en el análisis de datos conductuales. Dado que gran parte de este libro se ocupará de tales medidas —de hecho, hay toda una sección llamada "El análisis de varianza"— aquí solamente se caracterizarán promedios y varianzas. Los tres promedios principales (o medidas de tendencia central) usadas en investigación —media, mediana y moda— son resúmenes de los conjuntos de las medidas, a partir de los cuales se calculan. Los conjuntos de medidas son demasiado vastos y complejos para poder entenderlos de inmediato. Ellos están "representados" o resumidos por medidas de tendencia central. Éstas indican qué conjuntos de medidas "son parecidos" en promedio, pero también son comparados para probar relaciones. Más aún, las puntuaciones individuales pueden ser útiles, comparadas con aquellas que evalúan el estatus del individuo. Se dice, por ejemplo, que la puntuación individual de A está a tal y tal distancia de la media.

Mientras que la media es el promedio más usado en investigación y sus propiedades son tan deseables que justifican su posición preeminente, por otro lado, la mediana (la medida más medial de un conjunto de medidas) y la moda (la medida más frecuente) pueden algunas veces ser útiles en investigación. Por ejemplo, la mediana, además de ser una importante medida descriptiva, puede ayudarse en pruebas de significancia estadística donde la media es inapropiada (véase Bradley, 1968). El estudio de Allman, Walker, Hart, Laprade, Noel y Smith (1987), donde se comparó la efectividad y los efectos adversos de los colchones de aire y la terapia convencional en pacientes hospitalizados con úlceras de presión, sirve como un buen ejemplo del uso de la mediana como la medida primaria de tendencia central. La moda es usada principalmente para propósitos descriptivos, pero

puede ser útil en investigación para estudiar características de poblaciones y relaciones. Suponga que una prueba de aptitud matemática se aplicó a todos los aspirantes a ingresar a una universidad que apenas había abierto sus admisiones y que la distribución de las puntuaciones resultó ser bimodal. Suponga, además, que solamente se calculó una media y se comparó con las medias de los años anteriores, resultando ser considerablemente menor. La conclusión simple de que la aptitud matemática promedio de los aspirantes fue considerablemente menor que en años previos, oculta el hecho de que debido a la política abierta de admisión muchos aspirantes con antecedentes deficientes en matemáticas fueron admitidos. Aunque éste es un ejemplo obvio y escogido deliberadamente, debe observarse que el hecho de oscurecer fuentes importantes de diferencias puede ser más sutil. A menudo es útil en investigación calcular las medianas y las modas, así como las medias.[4]

Las principales medidas de variabilidad son la varianza y la desviación estándar. Éstas ya se han estudiado y se estudiarán en capítulos posteriores y sólo cabe decir que los reportes de investigación siempre deben incluir medidas de variabilidad. Las medias no deben reportarse sin desviaciones estándar (tampoco sin N, el tamaño de la muestra), ya que una adecuada interpretación de la investigación es virtualmente imposible sin los índices de variabilidad. Otra medida de variabilidad que en años recientes ha adquirido mayor importancia es el *rango*: la diferencia entre la medida más alta y la más baja de un conjunto de medidas. Ahora es posible, especialmente con muestras pequeñas (con N de 20, 15 o menos), usar el rango en pruebas de significancia estadística.

Medidas de relaciones

Hay muchas medidas útiles de relaciones: el coeficiente de correlación producto-momento (r), el coeficiente de correlación de rangos ordenados (*rho*), la razón de correlación (eta: η), la medida de distancia (D), el coeficiente phi (ϕ), el coeficiente de correlación múltiple (R), etcétera. Casi todos los coeficientes de correlación sin importar qué tan diferentes sean en derivación, apariencia, cálculo y uso, hacen en esencia lo mismo: expresan la extensión en que los pares de conjuntos de pares ordenados varían concomitantemente; informan al investigador la magnitud y (generalmente) la dirección de la relación. El valor de algunos varía de –1.00 a +1.00 pasando por 0, donde –1.00 y 1.00 indican una asociación negativa y positiva perfecta respectivamente, y el 0 indica una relación no discernible.

Las medidas de relación son, comparativamente, índices directos de relaciones, en el sentido de que a partir de ellas se adquiere una idea directa del grado de covariación de las variables. El cuadrado del coeficiente de correlación producto-momento, por ejemplo, es un estimado directo de la cantidad de varianza compartida por las variables. Se puede decir, al menos de forma general, qué tan alta o qué tan baja es la relación. Esto contrasta con las medidas de significancia estadística que indican si una relación es o no "significativa" a un nivel específico de significancia. Idealmente, cualquier análisis de datos de investigación debe incluir ambas clases de índices: medidas de significancia de una relación y medidas de la magnitud de la relación.

Las medidas de relación, pero sobre todo los coeficientes de correlación producto-momento, son poco usuales en cuanto a que están sujetos a formas extensas y elaboradas de análisis, principalmente análisis de regresión múltiple y análisis factorial (que se revisarán en capítulos posteriores). Por lo tanto, son herramientas extremadamente útiles y poderosas para el investigador.

---

[4] Diversos tipos de medias y de otras medidas de tendencia central son excepcionalmente bien explicadas en el libro de Tate (1955), un viejo pero valioso documento. Él también da un buen número de ejemplos de distribuciones y gráficos de varios tipos.

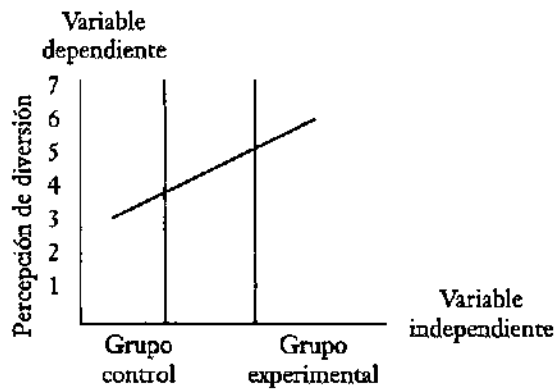
Análisis de diferencias

Los análisis de diferencias, particularmente el análisis de diferencias entre medias, ocupa una parte muy importante del análisis estadístico y de la inferencia. Es importante observar dos cosas acerca de los análisis de diferencias. Primero, por ningún motivo están confinados a las diferencias de medidas de tendencia central. Casi cualquier clase de diferencia puede ser analizada —entre frecuencias, proporciones, porcentajes, rangos, correlaciones y varianzas—. Consideremos las varianzas. Suponga que un psicólogo educativo desea averiguar si cierta forma de instrucción tiene el efecto de hacer a los alumnos más heterogéneos en el aprendizaje de conceptos. La diferencia entre las varianzas de los grupos que recibieron enseñanza por diferentes métodos puede ser probada fácilmente. También puede investigarse si grupos considerados homogéneos, lo son también en variables distintas a las usadas para formar los grupos (véase Comrey y Lee, 1995, pp. 229-234; Mattson, 1986, capítulo 10).

El segundo punto es más importante. Todos los análisis de diferencias son planeados con el propósito de estudiar relaciones. Suponga que alguien cree que el modificar la cantidad de narcisismo tendrá un efecto en las relaciones interpersonales. Carroll, Hoeningmann, Stovall y Whitehead (1996) crearon tres protocolos con diferentes niveles de narcisismo —extremo, moderado y ninguno— y después evaluaron el atractivo de los participantes hacia esa persona. La hipótesis de los investigadores se sustentó, ya que los participantes reportaron un mayor rechazo de la persona con un narcisismo extremo, que de aquellas con otros niveles de narcisismo, sin embargo lo que realmente interesa aquí no son estas diferencias, sino la relación entre el cambio de los niveles narcisismo y su efecto en cómo la gente percibe a la persona; entonces, las diferencias entre las medias realmente reflejan una relación entre la variable independiente y la variable dependiente. Si no hay diferencias significativas entre las medias, la correlación entre la variable independiente y la variable dependiente es 0; y, a la inversa, entre más grandes sean las diferencias, más alta será la correlación, siempre y cuando lo demás permanezca igual.

En el experimento de Strack, Martin y Stepper (1988), se estudió el efecto de la actividad facial de la gente en sus respuestas afectivas. Las personas en el grupo experimental recibieron instrucciones de sostener bolígrafos en la boca, usando solamente sus dientes, mientras veían caricaturas. Las personas del grupo control recibieron instrucciones de sostener los bolígrafos con sus labios, mientras veían el mismo estímulo. El grupo experi-

FIGURA 9.4



mental tuvo una media de 5.09 en una escala de 0 a 9 en la que evaluaban qué tan "divertida" era la caricatura. Sin embargo, el grupo control tuvo una media de 3.90. La diferencia es estadísticamente significativa, por lo que se puede concluir que hay una relación entre los músculos faciales utilizados (y aquellos no utilizados), y la diversión percibida. En capítulos anteriores se graficaron las relaciones entre variables medidas para mostrar la naturaleza de las relaciones. También es posible graficar la presente relación entre la variable independiente experimental (manipulada) y la variable dependiente. Esto se hizo en la figura 9.4, donde las medias se trazaron como se indica. Mientras que el trazo es más o menos arbitrario —por ejemplo, no hay unidades reales de línea base para la variable independiente—, la similitud con los gráficos previos es notoria y la idea básica de una relación ahora es clara.

Si el lector conserva siempre en mente que las relaciones son conjuntos de pares ordenados, la similitud conceptual de la figura 9.4 con gráficos anteriores será evidente. En los gráficos previos, cada miembro de cada par representa una puntuación. En la figura 9.4, un par ordenado consta de un tratamiento experimental y de una puntuación. Si se asigna el valor 1 al grupo experimental y 0 al grupo control, dos pares ordenados pueden ser los siguientes: (1, 5.09), (0, 3.90).

Análisis de varianza y métodos relacionados

Una buena parte de este libro está dedicada al análisis de varianza y los métodos relacionados con éste, por lo que no es necesario discutir mucho de esto por el momento. El lector necesita solamente tener en perspectiva este importante método de análisis. El análisis de varianza es lo que su nombre implica y más: un método para identificar, analizar y probar la significancia estadística de varianzas que provienen de diferentes fuentes de variación; es decir, que una variable dependiente tiene una cantidad total de varianza, parte de la cual es debida al tratamiento experimental, parte al error y parte a otras causas. El papel del análisis de varianza es trabajar con estas diferentes varianzas y fuentes de varianza. Estrictamente hablando, el análisis de varianza es más apropiado para datos experimentales que para datos no experimentales, aunque su inventor Fisher (1950), lo utilizó con ambos. Se considera, entonces, que el análisis de varianza es un método para el análisis de datos recabados en experimentos donde se utiliza, al menos, la aleatorización y manipulación de una variable independiente.

Probablemente no haya mejor forma de estudiar el diseño de investigación que a través del enfoque del análisis de varianza. Aquellos expertos en este enfoque, casi automáticamente piensan en modelos alternativos de análisis de varianza cuando se enfrentan a nuevos problemas de investigación. Tomemos el estudio de Rozin, Nemeroff, Wane y Sherrod (1989) sobre la ley del contagio. Esta ley establece que los objetos que han estado en contacto unos con otros, pueden continuar influenciándose entre sí a través de la transferencia de alguna de sus propiedades. Estos autores construyeron seis objetos (suéter, hamburguesa, manzana, cepillo de cabello [recibido], cepillo de cabello [dado] y un mechón de cabello) y los pusieron en contacto con cuatro diferentes personas (amigo, novio, alguien antipático y alguien indiferente). Se consideró a las cuatro diferentes personas como cuatro niveles de la variable categórica: fuente de origen. Se pidió a los participantes que evaluaran cada objeto en una escala de −100 a +100, donde −100 era "la cosa más desagradable que pueda usted imaginar" y +100 era "la cosa más agradable que usted pueda imaginar". Cero era el punto neutral. Rozin *et al.* (1989) analizaron los datos con un análisis varianza de un factor para cada objeto, usando la fuente de origen (diferentes personas que habían estado en contacto con el objeto) como la variable independiente. El análisis resultaría como el paradigma marcado (a) de la figura 9.5. Si los investigadores

FIGURA 9.5

(a)

Fuente de origen (personas en contacto con el objeto)			
Amigo	**Novio**	**Alguien antipático**	**Alguien indiferente**
Evaluaciones placenteras			

(b)

Objetos	Fuente de origen (personas en contacto con el objeto)			
	Amigo	**Novio**	**Alguien antipático**	**Alguien indiferente**
Suéter Hamburguesa Manzana Cepillo de cabello (recibido) Cepillo de cabello (dado) Mechón de cabello	Evaluaciones placenteras			

FIGURA 9.6

(a)

Proveedores de servicios de emergencia		
Médicos	**Enfermeras**	**Personal prehospitalario**
Conteo de bacteria estafilocóxica		

(b)

Turno laboral	Proveedores de servicios de emergencia		
	Médicos	**Enfermeras**	**Personal prehospitalario**
Día Noche	Conteo de bacteria estafilocóxica		

hubieran usado los objetos como otra variable independiente, pensando que los objetos afectan las evaluaciones de las personas, entonces el paradigma se vería como el marcado (b), que es una análisis de varianza de dos factores. Es claro que el análisis de varianza es un método importante para estudiar diferencias.

De la misma forma, un estudio de Jones, Hoerle y Riekse (1995) comparó la extensión de la presencia de la bacteria estafilococo en los estetoscopios de los proveedores de servicios de emergencia. Se utilizó un análisis de varianza de un factor para hacer esta comparación entre los médicos, enfermeras y personal prehospitalario. La variable independiente era el tipo de proveedor del servicio de emergencia y la variable dependiente era el conteo de bacteria estaflocóxica. La figura 9.6 (a) muestra el paradigma usado en este estudio. Si estos investigadores consideraran que podría haber una diferencia entre los proveedores de servicios de emergencia en los diferentes turnos de trabajo, el paradigma se expresaría como el que se encuentra en la figura 9.6 (b).

Análisis de perfiles

El *análisis de perfiles* es básicamente la evaluación de similitudes en los perfiles de individuos o grupos. Un *perfil* es un conjunto de medidas diferentes de un individuo o grupo, donde

cada una está expresada en la misma unidad de medida. Las puntuaciones de un individuo en un conjunto de diferentes pruebas constituye un perfil, siempre y cuando todas las puntuaciones hayan sido convertidas a un sistema común de medida, como percentiles, rangos y puntuaciones estándar. Los perfiles se han utilizado principalmente con propósitos diagnósticos —por ejemplo los perfiles de puntuaciones de una batería de pruebas son usados para evaluar y asesorar a alumnos de preparatoria—. Sin embargo, el análisis de perfiles ha incrementado su importancia en la investigación sociológica y psicológica, tal como se verá posteriormente cuando se estudie entre otras cosas, la metodología Q.

El análisis de perfiles tiene problemas especiales que requieren consideraciones cuidadosas por parte del investigador. La similitud, por ejemplo, no es una característica general de las personas; solamente hay similitud en características específicas o complejos de características (véase Cronbach y Gleser, 1953). Otra dificultad estriba en qué tipo de información se está dispuesto a sacrificar al calcular los índices de similitud de perfiles. Cuando se utiliza el coeficiente de correlación producto-momento —que es una medida de perfiles—, se pierde nivel, es decir, que se sacrifican las diferencias entre las medias. Esto es una pérdida de *elevación*. Las *r*s producto-momento sólo toman en cuenta la *forma*. Más aún, la *dispersión* (las diferencias en la variabilidad de los perfiles) se pierde al calcular otras clases de medidas de perfiles. En pocas palabras, la información puede perderse, y de hecho, se pierde. El estudiante encontrará una excelente ayuda y guía para el análisis de perfiles en el libro de Nunnally y Bernstein (1993) de psicometría, aunque el tratamiento no es elemental.

Análisis multivariado

Quizás las formas más importantes del análisis estadístico, especialmente en el estado actual del desarrollo de las ciencias del comportamiento, son los análisis multivariados y los análisis factoriales. *Análisis multivariado* es un término general usado para categorizar una familia de métodos analíticos cuya característica principal es el análisis simultáneo de k variables independientes y m variables dependientes. En este libro no nos preocuparemos demasiado acerca de la terminología usada en el análisis multivariado. Para algunos, el análisis multivariado incluye al análisis factorial y otras formas de análisis, como el análisis de regresión múltiple. *Multivariado*, para estas personas infiere más de una variable independiente o más de una variable dependiente, o *ambos*. Otros en el medio usan "análisis multivariado" solamente en el caso de que ambas, la variable dependiente y la variable independiente, sean múltiples. Si un análisis incluye, por ejemplo, cuatro variables independientes y dos variables dependientes, manejadas simultáneamente, entonces es un análisis multivariado.

Puede argumentarse que de todos los métodos de análisis, los métodos multivariados son los más poderosos y apropiados para la investigación científica del comportamiento. El argumento que apoya esta afirmación es muy amplio y complejo y nos apartaría del principal propósito. Básicamente descansa en la idea de que los problemas de investigación del comportamiento son, casi todos, de naturaleza multivariada y que no pueden ser resueltos con un enfoque bivariado (de dos variables), esto es, un enfoque que considere solamente una variable independiente y una variable dependiente a la vez. Esto ha quedado muy claro en mucha de la investigación educativa donde, por ejemplo, los determinantes del aprendizaje y aprovechamiento son complejos: inteligencia, motivación, clase social, instrucción, atmósfera escolar y del salón de clases, la organización escolar, etcétera. Evidentemente variables como éstas interactúan unas con otras y algunas veces unas contra otras, de maneras desconocidas, pero afectan el aprendizaje y el aprovechamiento. En otras palabras, para explicar los complejos fenómenos psicológicos o sociológicos de la

educación, se requiere de herramientas de diseño y análisis que sean capaces de manejar la complejidad que manifiestan, por sí mismas, las múltiples variables independientes y variables dependientes. Un argumento similar puede darse para la investigación psicológica y sociológica.

Este argumento y la realidad que subyace, imponen una pesada carga en aquellos individuos que enseñan y aprenden métodos y enfoques de investigación. Es poco realista e irresponsable estudiar y aprender solamente un enfoque que es básicamente bivariado en su concepción. Los métodos multivariados, sin embargo, son como la realidad conductual que tratan de reflejar: complejos y difíciles de entender. La necesidad pedagógica, en cuanto a este libro concierne, es tratar de expresar los fundamentos del pensamiento de investigación, diseño, métodos y análisis, principalmente a través de un enfoque bivariado modificado. Se extenderá este enfoque, tanto como sea posible, a los conceptos y métodos multivariados, esperando que el estudiante busque más adelante, después de haber recibido los fundamentos adecuados.

La *regresión múltiple*, que es probablemente la forma más útil de los métodos multivariados, analiza las influencias comunes y separadas de dos o más variables independientes sobre una variable dependiente. Esta afirmación tiene limitaciones, especialmente acerca de las contribuciones separadas de las variables independientes, lo que será discutido en el capítulo 33. El aumento del uso de la regresión múltiple como herramienta analítica de las ciencias del comportamiento se debe en su mayor parte, a las computadoras digitales de alta velocidad. Ezekiel y Fox (1959) son dos de los pocos autores cuyos libros sobre regresión múltiple, previa al gran uso de las computadoras, están disponibles. Ezekiel y Fox resumieron los estudios que utilizaron regresión múltiple, antes de la publicación de su libro en 1959. No había muchos de ellos. A partir de la gran disponibilidad de las computadoras y los programas de estadística, el número de estudios que usan la regresión múltiple se ha incrementado exponencialmente. Erlich y Lee (1978) propusieron un uso novedoso del análisis de regresión; lo utilizaron en puntuaciones de pruebas con el propósito de evaluar la responsabilidad educativa. Por otro lado, Griffiths, Bevil, O'Connor y Wieland (1995) usaron la regresión para predecir el nivel de competencia en un examen de anatomía y fisiología; entre las variables predictoras estaban el promedio, así como el tipo de escuela donde habían tomado los cursos propedéuticos de anatomía y fisiología. El método ha sido usado en cientos de estudios, probablemente por su flexibilidad, poder y aplicabilidad general a muchos tipos diferentes de problemas de investigación. (¡También tiene limitaciones!) Por eso, no puede ignorarse en este libro. Afortunadamente, no es tan difícil de entender y aprender a usarlo —dado el suficiente interés y deseo para hacerlo—.

La *correlación canónica* es una extensión lógica de la regresión múltiple. De hecho, es un método de regresión múltiple. Añade más de una variable dependiente al modelo de regresión múltiple; en otras palabras, maneja las relaciones entre conjuntos de variables independientes y conjuntos de variables dependientes por lo que es, teóricamente, un poderoso método de análisis. Sin embargo, tiene limitaciones que pueden restringir su utilidad, tales como la interpretación de los resultados que produce y su capacidad limitada para probar modelos teóricos.

El *análisis discriminante* también está estrechamente relacionado con la regresión múltiple. Como su nombre lo indica, su propósito es discriminar grupos entre sí con base en conjuntos de medidas. También es útil en asignar individuos a grupos, con base en sus puntuaciones en pruebas. Aunque esta explicación no es adecuada, será suficiente por ahora.

En esta etapa es difícil caracterizar, aun en un nivel superficial, la técnica conocida como *análisis multivariado de varianza* porque aún no se ha revisado el análisis de varianza. Por lo anterior se pospone su discusión.

El *análisis factorial* es esencialmente diferente en su clase y en su propósito de otros métodos multivariados. Su propósito fundamental es ayudar al investigador a descubrir e identificar las unidades o dimensiones llamadas *factores* que subyacen a muchas medidas. Por ahora se sabe, por ejemplo, que detrás de muchas medidas de habilidad e inteligencia subyacen algunas dimensiones generales o factores. La aptitud verbal y la aptitud matemática son dos de los factores más conocidos. Al estudiar actitudes sociales se han encontrado factores religiosos, económicos y educativos.

Los métodos multivariados mencionados anteriormente son "estándar" en el sentido de que a ellos nos referimos generalmente al usar el término "métodos multivariados". Sin embargo, hay otros métodos multivariados de igual, e inclusive, mayor importancia. Como se dijo en el prefacio, en un libro de esta naturaleza no es posible dar una explicación técnica adecuada y correcta de todos los métodos multivariados. Por ejemplo aunque tienen una enorme importancia, el análisis estructural de covarianza y el análisis de modelos log-lineales pueden ser demasiado complejos y difíciles de describir y explicar de una forma adecuada y completa. Lo mismo sucede con el método de análisis multidimensional y con el análisis de ruta, que no pueden ser presentados adecuadamente. ¿Entonces, qué se va a hacer? Algunos de estos enfoques y procedimientos son tan poderosos e importantes —de hecho están revolucionando la investigación conductual— que un libro que los ignore será un texto deficiente. La solución al problema fue también expuesto en el prefacio, y vale la pena repetirlo. Los enfoques más comunes y accesibles —el análisis de varianza, la regresión múltiple y al análisis factorial— serán presentados con suficientes detalles técnicos para permitir a un estudiante motivado y entusiasta aplicarlos e interpretar sus resultados. Otros métodos más complicados (como el análisis estructural de covarianza y los modelos log-lineales) serán descritos y explicados "conceptualmente" en sus propósitos y razonamientos con generosas citas y descripciones de investigaciones ficticias y reales. Tal enfoque será usado en capítulos posteriores con las siguientes tres metodologías.

El *análisis de ruta* es un método gráfico del estudio de las supuestas influencias directas e indirectas de las variables independientes entre sí y sobre las variables dependientes. En otras palabras, es un método para describir y probar "teorías" (véase Kerlinger y Pedhazur, 1973; Pedhazur, 1996). Quizás su principal virtud es que requiere que los investigadores expliciten el marco teórico de los problemas de investigación. Para lograr sus objetivos, el análisis de ruta usa los llamados diagramas causales o diagramas de ruta y el análisis de regresión. Los lectores pueden satisfacer un poco su curiosidad examinando, en el capítulo 34, uno o dos de los ejemplos del análisis de ruta que se dan ahí. El análisis de ruta ha sido un marco conceptual útil para explicar las relaciones entre variables. El autor acreditado en desarrollar el análisis de ruta fue Wright (1921). Las aplicaciones de Wright fueron en el campo de la genética. Duncan (1966) y Blalock (1971) popularizaron el trabajo de Wright en las ciencias del comportamiento. Es útil estudiar el análisis de ruta porque ayuda a entender más fácilmente el análisis estructural de covarianza. De hecho, el análisis de ruta forma parte del análisis estructural de covarianza, como se verá en un capítulo posterior.

El *análisis estructural de covarianza* —o modelamiento causal,[5] o modelos de ecuaciones estructurales— es el último enfoque del análisis de estructuras complejas de datos. Este método implica, principalmente, el análisis de variación conjunta de variables que están en una estructura dictada por la teoría. Por ejemplo, se puede estudiar la adecuación de las teorías de inteligencia mencionadas en capítulos anteriores, ajustando las teorías al marco del análisis estructural de covarianza para después evaluar qué tan bien pueden explicar los

[5] El término "modelamiento causal" ya no está en voga, pues ciertos estadistas prominentes señalaron que los análisis basados en la correlación, en las ciencias sociales y conductuales, no podían establecer causa y efecto.

datos reales de una prueba de inteligencia. El método —o más bien, la metodología— es una síntesis matemática y estadística ingeniosa del análisis factorial, la regresión múltiple, el análisis de ruta y la medición psicológica en un único sistema exhaustivo que puede expresar y probar formulaciones teóricas complejas de problemas de investigación. Su creación se atribuye a Joreskog (1970) y a sus asociados, aunque Bentler (1989) ha propuesto el método con fuerza, creando un algoritmo diferente (EQS) al de Joreskog (LISREL).

Los *modelos log-lineales* representan el método multivariado más reciente (o metodología) para analizar datos de frecuencia. Los métodos multivariados mencionados anteriormente están orientados sobre todo a analizar datos obtenidos de medidas continuas: puntuaciones de pruebas, medidas de escalas de aptitudes y personalidad, medidas de variables ecológicas y otros aspectos similares. Como se verá en el siguiente capítulo, los datos de la investigación conductual aparecen ocasionalmente como frecuencias, en especial de individuos, por ejemplo, número de hombres y mujeres, de minorías étnicas y minorías no étnicas, maestros y no-maestros, individuos de clase media y clase trabajadora; y católicos, protestantes y musulmanes. El análisis log-lineal hace posible estudiar combinaciones complejas de dichas variables nominales y, como el análisis estructural de covarianza, probar teorías de las relaciones e influencias de tales variables entre sí. Brevemente se caracterizará la metodología en un capítulo posterior, aunque el reducido espacio y las dificultades técnicas forzarán a limitar la explicación a las ideas básicas involucradas. Se verá, al menos, que como el análisis estructural de covarianza, es uno de los desarrollos metodológicos más poderosos e importantes de la última parte del siglo XX.

Índices

Un *índice* puede definirse de dos formas. Primero, un índice es un fenómeno observable que es sustituido por un fenómeno menos observable. Un termómetro, por ejemplo, da lectura de números que representan grados de temperatura; los números en un velocímetro indican a cuántos kilómetros por hora está avanzando un vehículo. Las puntuaciones de una prueba indican niveles de aprovechamiento, aptitudes verbales, grados de ansiedad, etcétera.

Una segunda, y quizás más útil definición para el investigador dice que *un índice es un número que está compuesto de dos o más números.* Un investigador realiza una serie de observaciones y deriva un solo número de las medidas de las observaciones para resumirlas y expresarlas en forma sucinta. Con esta definición, todas las sumas y promedios son índices ya que incluyen en una sola medida más de una medida. Pero la definición también incluye la idea de los índices como compuestos de diferentes medidas. Los coeficientes de correlación son de este tipo, ya que combinan diferentes medidas en una sola o en un índice.

Existen índices de clase social. Por ejemplo, se puede combinar el ingreso, la ocupación y el lugar de residencia para obtener un buen índice de la clase social. Un índice de cohesión puede obtenerse preguntando a los miembros de un grupo si les gustaría o no seguir formando parte del grupo. Sus respuestas pueden combinarse en un solo número. En negocios y en economía, el poder de compra del dólar americano varía en el tiempo, de manera que es necesario ajustar otros valores para hacer comparaciones significativas. Por ejemplo, tomemos la comparación del costo de un automóvil en 1997 respecto a su costo en 1950. Uno de los primeros pasos es determinar el poder de compra del dólar americano en 1997 y compararlo con el poder de compra que tenía en 1950. El "Bureau of Labor Statistics" (Oficina de estadísticas laborales) registra y publica regularmente el índice de precios al consumidor (IPC). Este índice se utiliza como una medida del costo de la vida.

Los índices son importantes en investigación porque simplifican las comparaciones. De hecho, permiten al investigador hacer comparaciones que, de otra forma, no podrían hacerse o que solamente podrían hacerse con mucha dificultad. Los datos brutos son generalmente demasiado complejos para entenderlos y manejarlos matemática y estadísticamente, por lo que deben reducirse a una forma manipulable. El porcentaje es un buen ejemplo, ya que transforma puntajes brutos a formas comparables.

Los índices generalmente toman la forma de cocientes: un número es dividido entre otro número. Los índices más útiles varían entre 0 y 1.00 o entre –1.00 y +1.00 pasando por 0. Esto los hace independientes del número de casos y permite hacer comparaciones entre muestras y entre estudios. (Éstos generalmente se expresan en forma decimal.) Hay dos tipos de cocientes: las tasas y las proporciones. Un tercer tipo es el porcentaje, que es una variante de la proporción.

Una *tasa* (o *razón*) es un compuesto de dos números que relaciona un número con otro en forma de fracción o de decimal. Cualquier fracción y cualquier cociente son una razón. Tanto el numerador, como el denominador (o ambos) de una razón pueden, por sí mismos, ser tasas. El propósito principal y utilidad de una razón es el relacional, ya que permite la comparación de números. Para hacer esto quizá sea mejor colocar el mayor de los dos números del cociente en el denominador. Esto satisface la condición mencionada anteriormente de tener el rango de los valores de la razón entre 0 y 1, o entre –1.00 y +1.00, pasando por 0. Sin embargo, esto no es absolutamente necesario. Suponga que se desea comparar la tasa de hombres y mujeres graduados de preparatoria con la tasa de hombres y mujeres graduados de secundaria, todo esto en un periodo de varios años. La tasa algunas veces será menor de 1.00 y otras veces será mayor de 1.00, ya que es posible que la preponderancia de un sexo sobre el otro cambie de un año a otro.

Algunas veces las razones dan información más precisa (en cierto sentido) que la que proveen las partes de las que están compuestas. Si se estudiara la relación entre variables educativas y la tasa de impuestos, por ejemplo, y se usaran las tasas de impuestos reales, podría obtenerse una idea errónea de dicha relación. Esto es porque las tasas de impuestos, por característica, son a menudo engañosas. Algunas comunidades con altas *tasas* de impuestos, en realidad tienen niveles relativamente bajos de impuestos. El gravamen de propiedad puede ser bajo. Para evitar las discrepancias entre una comunidad y otra, se puede calcular, para cada comunidad, la tasa de gravamen contra el gravamen real. Entonces una tasa de impuestos ajustada (una "verdadera" tasa de impuestos), puede calcularse multiplicando la tasa de impuestos en uso, por esta fracción. Esto producirá una cifra más precisa para ser usada en los cálculos de relaciones entre la tasa de impuestos y otras variables. La razón de probabilidad es un tipo de índice que es valioso cuando se consideran datos de frecuencia en tablas de contingencia. Se examinará este tipo de índice estadístico en mayor detalle cuando se explique el análisis de datos de frecuencia y el análisis loglineal.

Una *proporción* es una fracción donde el numerador es una de dos o más frecuencias observadas y el denominador la suma de las frecuencias observadas. La definición de probabilidad dada anteriormente, $p = s/(s + f)$, donde s es igual al número de éxitos y f es igual al número de fracasos, es una proporción. Considere dos números cualesquiera, 20 y 60. La razón de los dos números es 20/60 = .33. (También podría ser 60/20 = 3.) Si estos dos números fueran las frecuencias observadas de la presencia y de la ausencia de un atributo en una muestra total, donde N = 60 + 20 = 80, entonces la proporción sería: 20/(60 + 20) = .25. Otra proporción, por supuesto, es 60/80 = .75.

Un *porcentaje* es simplemente una proporción multiplicada por 100. En el ejemplo anterior sería 20/80 × 100 = 25%. El propósito principal de las proporciones y porcentajes es reducir diferentes conjuntos de números a conjuntos comparables de números con una

base común. Cualquier conjunto de frecuencias puede ser transformado a proporciones o porcentajes para facilitar la manipulación e interpretación estadística y sus manifestaciones.

Es necesario ser precavido ya que es frecuente que haya una mezcla de dos medidas falibles, donde los índices pueden ser peligrosos. El viejo método de calcular el CI (coeficiente intelectual) es un buen ejemplo. El numerador de la fracción es en sí mismo un índice, dado que la edad mental (EM), es un compuesto de varias medidas. Un ejemplo mejor es el llamado coeficiente de desempeño: CD = 100 × EE/EM, donde EE es la edad educacativa y EM es la edad mental. Tanto el numerador como el denominador de la fracción son índices complejos, ya que ambos son la mezcla de mediciones de confiabilidad variable. ¿Qué significa el índice resultante? ¿Cómo interpretarlo prudentemente? Es difícil de decir. En resumen, mientras que los índices son herramientas indispensables para el análisis científico, éstos deben ser usados con cuidado y precaución.

Indicadores sociales

Los indicadores forman una clase especial de variables, aunque están estrechamente relacionados con los índices —de hecho muchas veces son índices— de acuerdo a la definición anterior. Variables tales como el ingreso, expectativa de vida, fertilidad, calidad de vida, nivel de educación (de las personas) y ambiente, pueden ser llamados indicadores sociales. Es evidente que son variables porque es común que se realicen cálculos estadísticos con ellos. Los indicadores sociales son tanto variables como estadísticos. Antes de continuar, es necesario mencionar que Bauer (1966) realizó el primer trabajo con indicadores sociales. Desafortunadamente, es difícil definirlos y no se intentará hacerlo aquí de manera formal. El artículo de Jaeger (1978) documenta las dificultades para definir los indicadores sociales. Sin embargo, los lectores deben saber que la idea de indicadores sociales es importante y lo será más en el futuro. Su uso se está extendiendo a todos los campos y, eventualmente, será estudiado en forma sistemática desde un punto de vista científico, así como desde una perspectiva "pública" y social.

El interés de este libro son los indicadores sociales, entendidos como una clase de variables sociológicas y psicológicas que en el futuro pueden ser útiles para desarrollar y probar teorías científicas sobre las relaciones entre los fenómenos sociales y psicológicos. Ciertos indicadores sociales se usan ahora en los llamados estudios de modelamiento causal de desempeño educativo y ocupacional. En 1972 Duncan, Featherman y Duncan usaron la clase social, la ocupación de los padres y su ingreso, sólo por mencionar algunos. También se han utilizado indicadores psicosociales como la calidad de vida percibida o "felicidad". Un ejemplo de esto puede encontrarse en Campbell, Converse y Rodgers (1976). Sin embargo, en general, parece que se ha hecho poco trabajo metodológico sistemático para categorizar y estudiar los indicadores sociales, la relación entre ellos y sus relaciones con otras variables. La mayoría de los trabajos pueden ser considerados demográficos y estrechamente pragmáticos —en esencia descriptivos—. Sin embargo, una vez que los problemas de confiabilidad y validez son identificados y resueltos, este campo es sumamente prometedor, y deberá ofrecer a los científicos del comportamiento algo más que estadísticas como "en 1956 el 51.2% de la población eran mujeres", o "el 54% de la población mayor de 18 años tenía de 9 a 12 años de educación". Entre los estudios más prometedores se encuentran los realizados por los investigadores Vickie Mays y Susan Cochran acerca de los riesgos de las prácticas sexuales. En un estudio de Cochran DeLeeuw y Mays (1995), usaron dos métodos estadísticos —el análisis de homogeneidad y el análisis de rasgos latentes— para tener una óptima evaluación de los patrones de conducta

sexual. El uso efectivo de estos métodos reduce los indicadores múltiples a una única puntuación que puede ser usada como una variable de resultado en investigaciones relacionadas con el virus de la inmunodeficiencia humana (VIH). Con este tipo de investigación, es posible continuar buscando estudios de análisis factorial de indicadores y estudios de análisis de covarianza, donde los indicadores son variables de las estructuras analizadas. También se puede esperar un aumento general del uso de la idea de indicadores en las áreas sociales y psicológicas de investigación. Esto se ve fácilmente en investigación educativa donde el aprovechamiento de los niños parece estar afectado en formas complejas por diferentes clases de variables, algunas de las cuales son del género de los indicadores sociales. Una de las virtudes del movimiento de indicadores sociales es que dichas influencias sobre el aprovechamiento serán usadas de manera más consciente y sistemática para estudiar y probar teorías de aprovechamiento.

La interpretación de los datos de investigación

Al evaluar la investigación, los científicos pueden disentir en dos temas generales: los datos y la interpretación de los datos. Los desacuerdos sobre los datos se enfocan a problemas tales como la validez y confiabilidad de los instrumentos de medición y la adecuación del diseño de investigación, los métodos de observación y el análisis. Asumiendo competencia, los mayores desacuerdos generalmente se enfocan en la interpretación de los datos. La mayoría de los psicólogos, por ejemplo, estarán de acuerdo en los datos de los experimentos de reforzamiento, pero disentirán vigorosamente en la interpretación de los datos de los experimentos. Tales desacuerdos son, en parte, una función de la teoría. En un libro como éste no se puede profundizar en las interpretaciones de los diferentes puntos de vista teóricos, por lo que nos limitaremos a un objetivo, que es la aclaración de algunos preceptos comunes de la interpretación de los datos *dentro* de un estudio de investigación particular o de una serie de estudios.

Adecuación de los diseños de investigación, metodología, mediciones y análisis

Uno de los temas más importantes en este libro gira alrededor de qué tan apropiada es la metodología para el problema bajo investigación. El investigador generalmente tiene preferencia por ciertos diseños de investigación, métodos de observación, métodos de medición y tipos de análisis. Todos ellos deben ser congruentes y deben encajar unos con otros. Por ejemplo, no es adecuado utilizar un análisis propio de frecuencias con, digamos, medidas continuas tomadas de una escala de actitudes. Es muy importante que el diseño, los métodos de observación, las mediciones y el análisis estadístico sean apropiados para el problema de investigación.

El investigador debe examinar a fondo la adecuación técnica de los métodos, medidas y estadísticas. La adecuación de la interpretación de los datos depende de tal escrutinio. Por ejemplo, una fuente común de debilidad en la interpretación es la negligencia con los problemas de medición. Es una necesidad urgente poner particular atención a la confiabilidad y validez de las medidas de las variables, como se verá en capítulos posteriores. Aun las personas y organizaciones más capaces en investigación titubean en ocasiones. Por muchos años, por ejemplo, la medición de las actitudes sociales comúnmente llamadas "liberalismo" y "conservadurismo" han sido cuestionadas. Por un lado, se ha asumido —aun a la vista de evidencia contraria— que liberalismo y conservadurismo forman parte

de un mismo continuo. Por otro lado, las actitudes sociales han sido medidas con muy pocos reactivos. Incluso algunas organizaciones competentes, instituciones e individuos altamente respetables han cometido estos errores (Barber, 1976). No es un pecado grave equivocarse, pero el pecado real es sacar conclusiones precipitadas, como las características de las personas, con base en mediciones de confiabilidad y validez dudosas (véase Dawes, 1994).

El aceptar sin cuestionamiento la confiabilidad y validez de las mediciones de variables es un error grave. Los investigadores deben de ser especialmente cuidadosos al cuestionar la validez de sus mediciones, dado que todo el marco de la interpretación puede colapsarse sólo en este punto. Si un problema psicológico incluye la variable ansiedad, por ejemplo, y el análisis estadístico muestra una relación positiva entre la ansiedad y el logro, el investigador debe preguntarse a sí mismo y a los datos si la ansiedad medida (o manipulada) es el tipo de ansiedad propia del problema. El investigador puede, por ejemplo, haber medido la ansiedad cuando la variable problema era realmente una ansiedad general. De igual forma, debe preguntarse si la medida elegida de desempeño es válida para los propósitos de la investigación. Si el problema de investigación demanda la aplicación de principios, pero la medida del desempeño es una prueba estandarizada que enfatiza el conocimiento de hechos, entonces la interpretación de los datos puede ser errónea.

En otras palabras, nos enfrentamos aquí al hecho obvio, pero fácilmente ignorado, de que la adecuación de la interpretación depende de cada eslabón en la cadena metodológica, así como de lo apropiado que sea cada eslabón en el problema de investigación y la congruencia de los eslabones entre sí. Esto se ve claramente al revisar resultados negativos o no concluyentes.

Resultados negativos y no concluyentes

Los resultados negativos o no concluyentes son mucho más difíciles de interpretar que los resultados positivos. Cuando los resultados son positivos y cuando apoyan la hipótesis, uno interpreta los datos a través de las líneas de la teoría y del razonamiento que subyacen a las hipótesis. Aunque se formulen cuidadosamente preguntas críticas, las predicciones sostenidas son evidencia para la validez del razonamiento que está detrás del problema enunciado.

Ésta es una de las grandes virtudes de la predicción científica. Cuando se predice algo, y se planea y ejecuta un esquema para probar la predicción, y las cosas resultan como se predijo, lo adecuado del razonamiento y de la ejecución parece sustentarse, aunque nunca se puede estar completamente seguro. Los resultados, aunque predichos, pueden ser como son por razones muy diferentes a las que se creía. Más aún, el hecho de que toda la cadena compleja de teoría, las deducciones a partir de la teoría, el diseño, la metodología, las mediciones y el análisis ha llevado a un resultado presupuesto, es fuerte evidencia de que toda la estructura ha sido adecuada. Aquí, se hace una apuesta compleja, con la suerte en contra. Entonces se lanzan los dados de la investigación o se gira la ruleta de la investigación; si resulta el número predicho, el razonamiento y el procedimiento que llevaron a una predicción exitosa, parecerán ser adecuados. Si se puede repetir la hazaña, entonces la evidencia de lo adecuado de la predicción será aun más convincente.

Pero ahora tomemos el caso negativo. ¿Por qué fueron negativos los resultados? ¿Por qué no salieron como se predijo? Observe que cualquier eslabón débil en la cadena de una investigación puede causar resultados negativos. Esto puede deberse a una, varias o todas las siguientes causas: teoría e hipótesis incorrectas, metodología inapropiada o incorrecta, mediciones inadecuadas o pobres y análisis defectuosos. En 1976, Barber afirmó que in-

cluso podía ser el resultado de un planteamiento incorrecto. Todas estas causas deben ser examinadas y evaluadas minuciosamente para ver si los resultados negativos dependen de una, de varias o de todas ellas. Si se está completamente seguro de que la metodología, la medición y el análisis son adecuados, entonces los resultados negativos podrán ser una contribución definitiva al avance científico. Es con este tipo de resultados, que se puede tener cierta certeza de que las hipótesis son incorrectas.

Relaciones no hipotetizadas y hallazgos no anticipados

Probar relaciones hipotetizadas es algo que se enfatiza en este libro. Sin embargo, esto no significa que otras relaciones en los datos no sean buscadas y probadas; muy al contrario. En la práctica, los investigadores siempre están ansiosos por buscar y estudiar relaciones en sus datos. Las relaciones no predichas pueden ser una clave importante para un entendimiento más profundo de la teoría; pueden resaltar aspectos del problema que no se anticiparon cuando éste se formuló. Por lo tanto, los investigadores —al enfatizar relaciones hipotetizadas— siempre deben estar alertas de relaciones no anticipadas en sus datos.

Suponga que se hipotetiza que un agrupamiento homogéneo de los alumnos será benéfico para los alumnos brillantes pero no para los alumnos con menos habilidades; imagine que esta hipótesis se sustenta, pero se nota una diferencia aparente entre las áreas rurales y suburbanas. La relación parece más fuerte en las áreas suburbanas, ¡pero se encuentra invertida en algunas áreas rurales! Se analizan los datos usando la variable suburbano-rural y se encuentra que el agrupamiento homogéneo parece tener una influencia marcada en los niños brillantes en el área suburbana, pero que hay poca o ninguna influencia en el área rural. Éste sería un hallazgo verdaderamente importante.

Uno de los hallazgos más fuertes y mejor apoyados de la psicología moderna es que el reforzamiento positivo fortalece la tendencia a responder (véase Hergenhahn, 1996). Por ejemplo, se ha creído que para mejorar el aprendizaje de los niños, sus repuestas correctas a problemas deben ser reforzadas positivamente. Sin embargo, sorpresivamente se ha encontrado que la motivación externa a veces tiene efectos perjudiciales. El trabajo de Lepper, Greene y Nisbett (1973) demostró que el reforzamiento positivo extrínseco minaba el interés intrínseco de los niños en un actividad de dibujo, un resultado ciertamente no predecible a partir de la teoría del reforzamiento.[6]

Los hallazgos no predichos e inesperados deben ser tratados con mayor suspicacia que aquellos predichos y esperados. Antes de ser aceptados, deben ser probados en una investigación independientemente, en la que sean predichos y probados de manera específica. Sólo cuando una relación es probada deliberada y sistemáticamente, con los controles necesarios construidos en el diseño, se puede tener fe en ellos. Los hallazgos no anticipados pueden ser fortuitos o espurios.

Tukey (1977) desarrolló métodos para analizar datos en una investigación. El uso de estos métodos se llama *análisis exploratorio de datos*. Tukey, así como Hoaglin, Mosteller y Tukey (1985) han presentado varios diagramas de fácil construcción que resumen y describen los datos. Estos diagramas pueden proveer información útil al investigador para consideraciones adicionales. Uno de los más populares es el *diagrama de tallo y hojas*. Este diagrama es similar al histograma pero tiene la ventaja de no perder los datos originales. El método de tallo y hojas trabaja mejor cuando el tamaño de la muestra es menor de 100. El principio detrás de este método es que un tallo y una hoja se usan para representar

[6] El trabajo Lepper *et al.*, así como el de otros autores, sobre motivación intrínseca y extrínseca se revisa en los artículos de Cameron y Pierce (1994, 1996).

TABLA 9.1 *Datos ficticios usados para demostrar el método de tallo y hojas (N = 46)*

81	54	91	74	88	78	90	77	88	90	69	94	74	76	96	50	93	93	70	77	58	60	75
53	81	73	66	86	81	64	77	56	71	71	56	53	83	85	70	71	76	80	87	62	57	73

cada punto o valor. El tallo se coloca a la izquierda de una línea vertical y la hoja a la derecha de dicha línea.

Tomemos por ejemplo, los datos presentados en la tabla 9.1. La hoja para cada puntuación es el último dígito y el tallo lo constituyen el resto de dígitos de un número. Por ejemplo, el número 81 de la tabla 9.1 se vería como sigue:

Tallo	Hoja
8	1

La figura 9.7 muestra el diagrama final de tallo y hoja, al incluir todos los datos de la tabla 9.1. Con este diagrama se puede tener una idea clara de cómo se ve la distribución, ya que provee una descripción más detallada de los datos que las distribuciones ordinarias de frecuencia o los histogramas. El desarrollo de tales métodos puede ayudar a los investigadores a generar hipótesis para ser probadas.

Prueba, probabilidad e interpretación

La interpretación de los datos culmina en enunciados de probabilidad condicional del tipo "si p, entonces q". Estos enunciados se enriquecen al ser especificados como sigue: si p, entonces q, bajo las condiciones r, s y t. Generalmente se evitan los enunciados causales, por la conciencia de que no pueden ser realizados sin fuerte riesgo de error.

Quizás el problema de la prueba sea de mayor importancia práctica para el investigador que interpreta datos. Hay que aclarar que nada puede ser "probado" científicamente. Todo lo que puede hacerse es buscar evidencia para sostener que determinada proposición es cierta. Una prueba es un asunto deductivo. Los métodos experimentales de investigación no son métodos de prueba, son métodos controlados que brindan evidencia para apoyar la probable verdad o falsedad de las relaciones propuestas. En pocas palabras, ninguna investigación científica puede probar nada, por lo que la interpretación del análisis de los datos de investigación nunca debe usar la palabra *prueba*.

Afortunadamente, para propósitos prácticos de investigación, no es necesario preocuparse mucho acerca de la causalidad y de la prueba. La evidencia con niveles satisfactorios

FIGURA 9.7

Tallo	Hoja
5	0 3 3 4 6 6 7 8
6	0 2 4 6 9
7	0 0 1 1 1 3 3 4 4 5 6 6 7 7 7 8
8	0 1 1 1 3 5 6 7 8 8
9	0 0 1 3 3 4 6

de probabilidad es suficiente para el progreso científico. La causalidad y las pruebas fueron analizadas en este capítulo para sensibilizar al lector del peligro del uso impreciso de los términos. El entendimiento del razonamiento científico, y la práctica y el cuidado razonable en la interpretación de los datos de investigación, son útiles guardianes contra la inferencia inadecuada de datos a conclusiones, aun cuando no garanticen la validez de las interpretaciones.

RESUMEN DEL CAPÍTULO

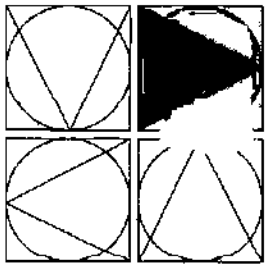
1. El análisis es el proceso de categorizar, ordenar, manipular y resumir datos para responder a las preguntas de investigación.
2. El propósito del análisis es reducir datos a una forma interpretativa, de manera que las relaciones puedan ser estudiadas y probadas.
3. La interpretación toma los resultados del análisis y hace inferencias y discute relaciones.
4. Los datos vienen en forma de medidas de frecuencia y medidas continuas.
5. El primer paso en cualquier análisis es la categorización o partición.
6. Los tipos de análisis estadísticos son:
   a) gráficos
   b) medidas de tendencia central y de variabilidad
   c) medidas de relaciones
   d) análisis de diferencias
   e) análisis de varianza
   f) análisis de perfiles y análisis multivariado
7. Los índices se usan para simplificar las comparaciones. Ejemplo de índices son los porcentajes, los cocientes y las tasas o razones.
8. Los datos y la interpretación de los datos son dos áreas en las que los científicos disienten.
9. Cuando se interpretan los datos de investigación, se debe considerar la adecuación técnica de la metodología de investigación, de los procesos de medición y de la estadística utilizados.
10. Los resultados negativos o no concluyentes son mucho más difíciles de interpretar que los resultados positivos.
11. Al conducir un estudio de investigación, pueden surgir relaciones no hipotetizadas y hallazgos no anticipados.
12. Los hallazgos no predichos e inesperados deben de ser tratados con más suspicacia que los hallazgos predichos y esperados.

SUGERENCIAS DE ESTUDIO

1. Suponga que usted desea estudiar la relación entre la clase social y el resultado de una prueba de ansiedad. ¿Cuáles son las dos principales posibilidades para analizar los datos? (omita la posibilidad de calcular un coeficiente de correlación). Establezca dos estructuras analíticas.
2. Suponga que usted quiere agregar la variable sexo al problema anterior. Establezca las dos clases de paradigmas analíticos.
3. Suponga que un investigador ha probado los efectos de tres métodos de enseñanza de lectura, en la ejecución de la lectura. Tenía 30 sujetos en cada grupo y una pun-

tuación de la ejecución de lectura para cada sujeto. También incluyó el género como una variable independiente: la mitad de los sujetos eran varones y la otra mitad eran mujeres. ¿Cómo se vería su paradigma analítico? ¿Qué va dentro de las casillas?

4. Estudie la figura 9.3. ¿Representan estos diseños o paradigmas de análisis de varianza la partición de las variables? ¿Por qué sí? ¿por qué no? ¿Por qué es importante la partición al establecer diseños de investigación y al analizar los datos? ¿Tienen las reglas de categorización (y partición) algún efecto en la interpretación de los datos? Si es así, ¿qué efectos pueden tener? (Considere los efectos de violar las dos reglas básicas de partición.)



CAPÍTULO 10

EL ANÁLISIS DE FRECUENCIAS

- TERMINOLOGÍA DE DATOS Y VARIABLES
- TABULACIÓN CRUZADA: DEFINICIONES Y PROPÓSITO
- TABULACIÓN CRUZADA SIMPLE Y REGLAS PARA LA CONSTRUCCIÓN DE UNA TABULACIÓN CRUZADA
- CÁLCULO DE PORCENTAJES
- SIGNIFICANCIA ESTADÍSTICA Y LA PRUEBA χ^2
- NIVELES DE SIGNIFICANCIA ESTADÍSTICA
- TIPOS DE TABLAS CRUZADAS Y TABLAS
  - **Tablas unidimensionales**
  - **Tablas bidimensionales**
  - **Tablas bidimensionales, dicotomías "verdaderas" y medidas continuas**
  - **Tablas de tres dimensiones y de *k*-dimensiones**
- ESPECIFICACIÓN
- TABULACIÓN CRUZADA, RELACIONES Y PARES ORDENADOS
  - **La razón de probabilidad**
  - **Análisis multivariado de datos de frecuencia**
  - **Anexo computacional**

Hasta ahora se ha hablado principalmente *acerca* del análisis. Ahora se explicará cómo *hacer* el análisis. La forma más simple de analizar datos para estudiar relaciones es por medio de la partición cruzada de frecuencias. Como se aprendió en el capítulo 4, la partición cruzada es una nueva partición del conjunto U, para formar los subconjuntos de la forma $A \cap B$; es decir, que se forman subconjuntos de la forma $A \cap B$ de los subconjuntos conocidos A y B de U. Se dieron ejemplos en el capítulo 4 y se darán más en breve. La expresión "partición cruzada" se refiere a un proceso abstracto de la teoría de conjuntos. Sin embargo, ahora que la idea de la partición cruzada se aplique al análisis de frecuencias para estudiar relaciones entre variables, se le llamará *tabulación cruzada*, aunque también se le ha llamado algunas veces *fracción cruzada*. El análisis que se mostrará también es conocido como análisis de *contingencia* o análisis de tabla de contingencia.

TABLA 10.1 *Relación entre afiliación al partido político y el voto de conciliación presupuestal, en el senado de Estados Unidos, 1995*

	Republicano	Demócrata	
En contra	1	46	47
	2%	100%	
A favor	52	0	52
	98%	0%	
	53	46	99

Fuente: Datos del *Congressional Quarterly* (1996).

Dado que no es posible seguir adelante sin la estadística, se introducirá una forma de análisis estadístico comúnmente asociado con las frecuencias, la prueba de χ^2 (chi cuadrada), y el concepto de "significancia" estadística. Este estudio de la tabulación cruzada y la χ^2 servirá como introducción a la estadística.

La pugna política entre republicanos y demócratas a menudo se refleja dramáticamente en los votos del Congreso. Una de las recientes votaciones importantes en el senado de Estados Unidos se llevó a cabo en el proyecto de ley fiscal de 1996, referente a la conciliación presupuestal. La lucha republicano-demócrata durante el invierno de 1995 se centró en las propuestas de balance del presupuesto hacia el año 2002: los republicanos se manifestaban, generalmente, a favor de las propuestas y los demócratas en contra de ellas, incluyendo al presidente Clinton. Una de estas propuestas era reducir los gastos en servicios de asistencia social y reducir los impuestos. El proyecto de ley se aprobó por 52 a 47. Esto fue una derrota para el presidente. Lo interesante aquí son los resultados de los votos republicano-demócratas, que se muestran en la tabla 10.1. Por las frecuencias (en este caso) es claro que hay una fuerte relación entre la afiliación al partido político y el voto en la propuesta de ley sobre el presupuesto: los demócratas votaron en contra y los republicanos votaron a favor.

No todas las frecuencias en la tabulación cruzada son así de claras. En la práctica es común calcular porcentajes. Si se hace en una forma que será descrita posteriormente, los porcentajes son los presentados en la esquina inferior derecha de cada casilla. Se nota la fuerza de la relación entre la afiliación al partido político y el voto: 98% de los republicanos votaron a favor y 100% de los demócratas votaron en contra.

Estudios de votaciones similares en la misma época muestran la misma relación general. Por ejemplo, los votos para imponer sanciones a los médicos que realizan abortos tardíos se pueden observar en la tabla 10.2. Nuevamente la relación es fuerte, aunque no tanto como en la votación para la conciliación presupuestal (observe los porcentajes). Un voto "en contra" en este proyecto de ley apoya la posición del presidente.

TABLA 10.2 *Voto del senado de Estados Unidos para imponer sanciones a los médicos que realizan abortos tardíos, 1995*

	Republicano	Demócrata	
En contra	45	9	54
	85%	20%	
A favor	8	36	44
	15%	80%	
	53	45	98