

CAPITULO UNO

El papel y la importancia de la investigación

En Neil J. Salkind "Métodos de Investigación", Prentice Hall, México, 1998.

Lo que aprenderá en este capítulo

- Quién hace investigación y por qué
- Cómo se define la investigación y cuáles son sus fines principales
- Qué es un modelo de investigación científica y cómo guía las actividades de indagación
- Lo que es y lo que no es la investigación
- El quehacer de los investigadores y la manera en que lo llevan a cabo
- Las características de una buena investigación
- Cómo un método de investigación científica guía las actividades de indagación
- Los diferentes tipos de métodos de investigación y ejemplos

Presentamos... la Investigación

Camine por un pasillo de cualquier edificio de su universidad donde tengan sus oficinas profesores de ciencias sociales, de departamentos como psicología, sociología y desarrollo humano. ¿Ve usted hombres barbudos, desaliñados, con batas blancas y pantalones arrugados, fumando pipa y encorvados sobre el teclado de su computadora mascullando ininteligiblemente?

¿Investigadores en plena actividad? No. ¿Estereotipos de cómo se ven y qué hacen los científicos? Sí. Lo que seguramente verá en los pasillos de su edificio de aulas o en las oficinas de los profesores es hombres y mujeres de todas edades que trabajan intensamente, decididos a encontrar la respuesta a un elemento más del gran acertijo que nos ayuda a entender el comportamiento humano un poquito mejor que la generación anterior de científicos.

Igual que todo mundo, estas personas acuden al trabajo en la mañana, pero a diferencia de mucha otra gente estos investigadores tienen una pasión inextinguible por entender lo que estudian y por acercarse lo más posible a la "verdad". Aunque tales verdades son elusivas y a veces hasta inalcanzables, los investigadores encaminan sus trabajos hacia ellas por la satisfacción de contestar preguntas importantes y luego utilizar esta nueva información para ayudar a sus semejantes. Programas de intervención temprana, tratamientos de psicopatología, técnicas de solución de conflictos, programas antidrogas eficaces e incluso cambios en las políticas y las leyes respecto a cuándo la gente debe jubilarse, han sido resultado de las evidencias que investigadores, entre los que podrían estar sus profesores, han obtenido. Si bien no siempre es perfecta, cada bloque de información de un nuevo estudio o una idea nueva para un estudio contribuye a un vasto legado de conocimientos para la siguiente generación de investigadores, entre los que seguramente usted estará.

Es posible que ya conozca y aprecie algunos aspectos del mundo de la investigación. El propósito de este libro es proporcionarle los medios que necesita para hacer todavía más, como por ejemplo:

- entender a fondo el proceso de investigación,
- prepararse para realizar investigaciones propias,
- aprender a juzgar la calidad de una investigación,
- aprender a leer, explorar y resumir otras investigaciones,
- averiguar de qué se trata Internet y cómo puede utilizarse en las actividades de investigación cotidianas,
- descubrir los misterios de las estadísticas básicas, y darse cuenta de lo fáciles que pueden ser,
- medir los comportamientos, rasgos o atributos que le interesan,
- recopilar los tipos de datos relacionados con su área de interés,
- utilizar uno de los principales paquetes de estadística (Paquete Estadístico para las Ciencias Sociales, SPSS) para analizar datos,
- diseñar estudios de investigación que contesten la pregunta cuya respuesta usted busca, y
- escribir el tipo de propuesta de investigación que le permitirá asumir el control, aquella que demuestre que usted domina el contenido de la investigación y la forma como debe llevarse a cabo.
- ¿Suena ambicioso? ¿Un poco aterrador? ¿Emocionante? Tal vez todo eso y más, pero lo que sí no es una labor de investigación aburrida. Esto es cierto sobre todo si consideramos que el trabajo que podría realizar en esta clase, así como la propuesta de investigación que podría escribir, podría ser la clave para expandir nuestro conocimiento y entendimiento del comportamiento humano.

Ahí está usted, entonces; a punto de iniciar lo que probablemente sea su primer curso en el área de métodos de investigación y preguntándose todo tipo de cosas, desde qué hacen los investigadores hasta qué tema escogerá para su tesis. Tranquilo. Miles de estudiantes han estado aquí antes que usted, y casi todos han salido con un conocimiento funcional de qué es la **investigación**, cómo se hace y qué distingue un buen proyecto de investigación de uno que está destinado al fracaso. Sujétese y... ¡vámonos!; el viaje será emocionante.

Lo que es y lo que no es la investigación

Tal vez lo mejor sea comenzar con una mirada a lo que los investigadores realmente hacen para ganarse la vida. Para ello, ¿por qué no escoger algunos de los mejores? Presentaremos algunos investigadores, los premios que han ganado y el enfoque de sus trabajos. Los distintos premios fueron otorgados por la American Psychological Association como reconocimiento de una labor sobresaliente. Toda esta gente comenzó en una clase como a la que usted está asistiendo, y leyó un libro similar al que tiene en sus manos. Su interés en la investigación y en un aspecto específico crecieron y crecieron hasta convertirse en la labor de su vida.

Leonard D. Eron recibió el Gold Medal Award por Contribuciones Vitales de un Psicólogo al Interés Público en 1995 por sus aportaciones a la teoría y la comprensión de la medición psicológica. En términos específicos, su trabajo se ha concentrado en describir y comprender el papel que desempeña la agresión de los niños y cómo prevenirla. Frances K. Graham recibió el Gold Medal por Logros Vitales en Psicología en 1995 por sus trabajos en psicofisiología y sus descubrimientos acerca del funcionamiento de la mente del niño recién nacido. Peter J. Lang recibió un premio de Contribución Científica Distinguida en 1995 por sus investigaciones sobre las emociones, sobre todo el temor y la ansiedad.

Lo que esta gente tiene en común es que, en algún momento de su carrera profesional, fueron participantes activos en el proceso de hacer investigación. La investigación es un proceso por el cual se descubren conocimientos nuevos. Una **teoría** de la motivación o el desarrollo, por ejemplo, nos ayuda a organizar información nueva en un cuerpo coherente, un conjunto de ideas relacionadas que explican sucesos que han ocurrido y predicen eventos que ocurrirán. Las teorías son una parte importante de la ciencia. No obstante, es en la planta baja, por así decirlo, que el investigador trabaja para poner en marcha el proceso, añadiendo unas cuantas ideas aquí y una que otra especulación nueva allá, hasta que todo se conjunta para formar un cuerpo de conocimientos.

La investigación de alta calidad se caracteriza por varios atributos distintos:

1. Se basa en el trabajo de otros.
2. Se puede repetir.
3. Se puede generalizar a otras situaciones.
4. Se basa en algún razonamiento lógico y está vinculado a una teoría.
5. ¡Se puede hacer!
6. Genera nuevas preguntas o es de naturaleza cíclica.
7. Es incremental.
8. Es una actividad apolítica que debe emprenderse con el fin de mejorar la sociedad.

En primer lugar, *la investigación es una actividad basada en el trabajo de otros investigadores*. Esto no quiere decir que debemos copiar el trabajo de otros (lo cual sería plagio), sino que siempre debemos examinar los trabajos que ya se han hecho para tener una base sobre la cual realizar nuestro propio trabajo. Por ejemplo, si se han realizado 200 estudios sobre la relación entre el número de niños en una familia y el nivel promedio de sus puntajes de IQ (coeficiente de inteligencia, CI), no debemos hacer caso omiso de esos estudios. Quizá no queramos repetir ninguno de esos estudios, pero ciertamente debemos tener en cuenta las metodologías que se utilizaron y los resultados que se obtuvieron al planear nuestras propias investigaciones en el área.

Un buen ejemplo de este principio es el enorme esfuerzo intelectual y científico que se invirtió en la creación de la bomba atómica. Aunque es un arma pavorosa, cientos de científicos de primer nivel de todo el mundo se organizaron en diferentes lugares en un esfuerzo intenso y altamente motivado por combinar sus conocimientos. Lo singular de este esfuerzo es que se realizó en tiempo comprimido; muchas personas que probablemente habrían compartido sus trabajos de todos modos lo hicieron en días en lugar de meses a causa de la urgencia militar y política de la época. Lo que se descubría un día se convertía literalmente en la base para los experimentos del día siguiente.

Segundo, ya que estamos hablando de otros estudios, *la investigación es una actividad que puede repetirse*. Si

alguien realiza un estudio de investigación que examina la relación entre la capacidad para resolver problemas y el talento musical, los resultados del experimento deberán ser repetibles por dos razones. Primera, uno de los rasgos distintivos de cualquier hallazgo científico plausible es que puede repetirse. Si podemos hilar oro a partir de paja, deberemos poder hacerlo siempre, ¿no? ¿Y qué tal un nuevo método para enseñar a los niños a leer? O desarrollar programas de intervención temprana que producen resultados similares cuando se repiten? Segunda, si los resultados de un experimento pueden repetirse, esto significa que pueden servir como base para investigaciones adicionales en la misma área.

Tercero, *la buena investigación se puede generalizar a otras situaciones*. Esto significa, por ejemplo, que si se averigua que los adolescentes son especialmente susceptibles a la presión de sus compañeros en una situación, los resultados probablemente seguirán siendo válidos en una situación diferente, pero relacionada con la anterior. Si bien la generalizabilidad de algunas investigaciones es limitada, porque es difícil duplicar las condiciones exactas en las que se efectuó la investigación, los resultados de casi todas las investigaciones pueden aportar al menos algo a otras situaciones.

Cuarto, *la investigación se basa en algún razonamiento lógico y esta vinculado a una teoría*. Las ideas de investigación no son autosuficientes, tan sólo como preguntas interesantes. Más bien, la actividad de investigación proporciona respuestas a preguntas que ayudan a llenar los huecos de lo que puede ser un rompecabezas grande y complicado. No es razonable esperar que alguien, con un solo magno proyecto de investigación, entienda cabalmente el proceso del desarrollo intelectual de los niños o la razón por la que los adolescentes forman grupos o lo que realmente les sucede a los adultos en una crisis alrededor de los 40 años. Todas esas áreas principales de investigación deben desglosarse en elementos más pequeños, y todos esos elementos deben ligarse con un tema común, que las más de las veces es alguna teoría subyacente que guía la investigación.

Quinto, y por supuesto, *¡la investigación debe poder hacerse!* Con demasiada frecuencia, sobre todo si el científico es joven o tiene poca experiencia (como usted), el reto de encontrar una idea factible presiona tanto que casi cualquier cosa se considera como tema de investigación. Así, los profesores suelen recibir propuestas de los estudiantes que dicen cosas como: "El propósito de la presente investigación es determinar si es posible reducir el uso de drogas mediante la exposición a anuncios comerciales por televisión". Este nivel de ambigüedad y falta de un marco conceptual hace que la propuesta sea casi inútil y ciertamente no factible. La buena investigación hace una pregunta que se puede contestar, y lo hace de forma oportuna.

Sexto, *la investigación genera nuevas preguntas y es de naturaleza cíclica*. Sí, "lo que da vueltas, regresa". Las respuestas a las preguntas de investigación de hoy serán la base para las preguntas de investigación que se harán mañana. Aprenderá más acerca de este proceso un poco más adelante en este capítulo cuando describamos un método de investigación científica.

Séptimo, *la investigación es incremental* (es decir, que va creciendo con base en lo que ya se tiene). Ningún científico se yergue solo; todos se paran sobre los hombros de otros. Las contribuciones por lo regular se hacen en fragmentos pequeños, fáciles de definir. El primer estudio sobre el desarrollo del lenguaje no contestó todas las preguntas sobre la adquisición del lenguaje, ni el último estudio realizado fue el que colocó el betún sobre el pastel. Más bien, todos los estudios en un área dada se conjuntan para producir un cuerpo de conocimientos que diferentes investigadores comparten y que proporciona las bases para investigaciones subsecuentes. El todo, o todos los conocimientos acerca de un área en particular, es más que la suma de las partes, porque cada nuevo avance en la investigación no sólo nos informa, sino que también ayuda a colocar otros hallazgos en una perspectiva diferente y a menudo fructífera.

Por último, idealmente, *la investigación es una actividad apolítica que debe emprenderse con el fin de mejorar la sociedad*. Hago hincapié en *idealmente*, porque con demasiada frecuencia algún grupo de interés especial u otro dicta la forma como debe gastarse el financiamiento dedicado a la investigación. El hallazgo de una vacuna para el SIDA no debe depender de las actitudes que ciertos individuos tengan hacia la forma como cada quién vive su vida. Asimismo, que deban apoyarse o no los programas de intervención temprana debe ser independiente de las opiniones personales.

Sin dejar de ser apolítica, la investigación debe tener como meta final el mejoramiento de la sociedad. Los investigadores no mantienen a mujeres embarazadas sin comer para estudiar los efectos de la desnutrición sobre los niños. Con objeto de examinar el vínculo entre la tensión y la nutrición, los investigadores no obligan a sujetos adultos a ingerir dietas especiales que podrían ser dañinas. Estas prácticas poco éticas no se justificarían por el fin perseguido, sobre todo al haber otras formas de contestar tales preguntas sin recurrir a prácticas posiblemente

peijudiciales.

Si estos atributos caracterizan la buena investigación, ¿qué es una mala investigación? La mala investigación adopta el enfoque opuesto al que hemos delineado, y más. En síntesis, la mala investigación es el viaje de pesca que se emprende en busca de algo importante que simplemente no se encontrará. Es plagiar el trabajo de otras personas. Falsificar datos para demostrar algo. Falsear la información y engañar a los participantes. Por desgracia, hay investigadores cuyo trabajo se caracteriza por tales prácticas, pero son una minoría.

Un modelo de investigación científica

En los últimos 20 años, el público ha conocido las vicisitudes y tribulaciones del proceso de investigación descrito por cientos de libros acerca del trabajo cotidiano de científicos de todo el mundo. Algunos de los mejores de esos libros en inglés se listan en la sección *¿Quiere saber más?* al final de este capítulo. Tal vez usted pueda consultarlos visitando los sitios en la red de algunas bibliotecas estadounidenses.

Sea cual sea el contenido específico de tales historias, todas tienen algo en común. El trabajo se llevó a cabo ajustándose a pautas que permitieron a los investigadores ir del punto A al punto Z sin perder la confianza en que estaban en el camino correcto para encontrar (lo que esperaban fuera) una respuesta adecuada a las preguntas que habían planteado.

Sus métodos y conclusiones no fueron erráticas gracias a una práctica importante: todos ellos comparten la misma filosofía general acerca de cómo deben contestarse las preguntas sobre comportamiento humano. Además, para que los científicos puedan confiar en sus colegas, en el sentido de confiar en los resultados que sus estudios producen, deben tener algo en común aparte de las buenas intenciones. Lo que comparten es una secuencia estándar de pasos para formular y contestar una pregunta. Dicha secuencia recibe el nombre de método científico.

Cuando leemos en un artículo de una revista que el método A es más eficaz que el método B para mejorar la retención en la memoria, podemos estar seguros de que siguieron los pasos que describiremos a continuación, en una forma o en otra. Al existir un acuerdo en cuanto al método general empleado para contestar la pregunta, los resultados de esta comparación del método A y el método B se pueden aplicar al siguiente estudio, el cual tal vez investigará variaciones del método A y cómo y por qué funcionan. Los esfuerzos de investigación de los psicólogos del desarrollo, gerontólogos (especialistas en el envejecimiento), lingüistas y psicofisiólogos dependen de la integridad de este proceso.

He aquí un conjunto de tales pasos (que se muestra en la figura 1.1) como parte de un modelo de investigación científica. El objetivo de este modelo es encontrar "la verdad" (¡lo que sea que eso signifique!) o, en otras palabras, utilizar un método que produzca una respuesta razonable y sólida a preguntas importantes que enriquecerán nuestro entendimiento del comportamiento humano.

Utilizaremos un interesante y estimulante tema, los efectos de la televisión sobre los niños, como ejemplo de los distintos pasos del modelo.

Hacer la pregunta

Hacer nuevas preguntas

Identificar los factores importantes

Reconsiderar la teoría

Formular una hipótesis

Trabajar con la hipótesis

Recopilar la información pertinente

Probar la hipótesis

Figura 1.1 Pasos del proceso de investigación. Cada paso prepara el escenario para el siguiente.

Planteamiento de la pregunta

¿Recuerda el cuento del Mago de Oz? Cuando Dorothy se dio cuenta de que necesitaba llegar a la Ciudad Esmeralda, preguntó a Glinda, la bruja buena: "Pero, ¿por dónde empiezo?" La respuesta de Glinda: "Casi todo mundo comienza por el principio, querida" se puede aplicar a casi cualquier empresa científica.

Nuestro primer paso, y el más importante, es hacer una pregunta (*Me pregunto qué sucedería si...*) o identificar una necesidad (*Tenemos que encontrar una manera de...*) que surge como consecuencia de la curiosidad y que nos obliga a encontrar una respuesta. Por ejemplo, podría intrigarnos la forma en que ver televisión afecta el desarrollo de las habilidades de lenguaje de los niños. También podríamos sentir la necesidad de averiguar cuál sería la forma más eficaz de utilizar la televisión para educar a niños y adultos acerca de los peligros de las drogas.

Tales preguntas se expresan informalmente y a menudo se utilizan como fuente de discusión y estímulo acerca del rumbo que debe tomar el tema de investigación específico ¿De dónde salen esas preguntas? Casi nunca surgen de los confines de un aula o laboratorio. Más bien, las preguntas saltan (en el pleno sentido de la palabra) de nuestra imaginación y nuestras propias experiencias, enriquecidas por los mundos de la ciencia, el arte, la música y la literatura. No es coincidencia que muchas obras de ficción se basen en hechos (como la mal llamada ciencia ficción). El científico verdaderamente creativo siempre está pensando tanto en soluciones a interrogantes existentes como en la siguiente pregunta importante que hará. Cuando Louis Pasteur dijo que "el azar favorece a la mente preparada", lo que en realidad estaba diciendo es, aprovecha todas las experiencias que puedas, tanto en la escuela como fuera de ella. Sólo así estarás bien preparado para reconocer la importancia de ciertos sucesos que actuarán como estímulo para actividades de investigación más rigurosas.

Las preguntas pueden ser tan amplias como indagar acerca de los efectos de la televisión sobre el desarrollo del lenguaje, o tan específicas como la relación entre el contenido de ciertos anuncios de televisión y los hábitos de compra de los adolescentes. Sea cual sea su contenido y su profundidad, las preguntas son el primer paso en cualquier empresa científica.

Identificar los factores importantes

Una vez planteada la pregunta, el siguiente paso es **identificar los factores que habría que examinar para contestar la pregunta**. Los factores podrían ser muy sencillos, como la edad del niño o su condición socioeconómica, o medidas más complicadas de los efectos que los dibujos animados violentos tienen sobre el comportamiento del niño.

Demos una ojeada a esta lista de factores que han sido escudriñados en los últimos 10 años por diversos investigadores interesados en los efectos de la televisión sobre los niños:

- edad del niño
- grado de violencia en los programas
- etapa del desarrollo cognoscitivo del niño
- actitud del productor
- expresión facial
- toma de decisiones
- descripción de los patrones de ver televisión, hecha por la madre
- estimulación emocional
- diferencias étnicas en la respuesta a los programas de televisión
- patrones de comunicación familiar

Y éstos son sólo 10 de los cientos de temas que podrían explorarse, cada uno de los cuales define factores importantes. Pero, de todos los factores que podrían ser importantes y que nos ayudan a entender mejor los efectos de la televisión, ¿cuáles deberíamos escoger como foco? **En general, conviene seleccionar factores que**

- **no se hayan investigado antes.**

- contribuyan al entendimiento de la pregunta que se está planteando,
- sean susceptibles de investigarse,
- tengan algún interés personal o profesional para el investigador, y
- ¡conduzcan a otra pregunta!

Si ya de por sí es difícil definir la naturaleza del problema que se desea estudiar, no hablemos de generar preguntas que lleven a nuevas preguntas. Pero una vez iniciado el viaje que lo convertirá en un científico, usted se ha convertido en miembro de un grupo selecto que tiene la responsabilidad de contribuir a la bibliografía científica no sólo con sus acciones, sino también con lo que considera que debe hacerse.

Formulación de una hipótesis

Cuando a una niña de nueve años se le preguntó qué creía que era una hipótesis, dio la que quizá sea la mejor respuesta: una conjetura educada. Surge una hipótesis cuando las preguntas se transforman en declaraciones que expresan las relaciones entre variables como un enunciado del tipo "si... entonces "

Por ejemplo, si la pregunta es "¿Qué efectos tiene sobre los muchachos ver violencia en la televisión?", la hipótesis podría ser *Los muchachos que ven actos agresivos en los dibujos animados durante las horas de mayor audiencia tienen una probabilidad más alta de exhibir comportamientos agresivos inmediatamente después de la sesión de televisión, que los muchachos que ven actos no agresivos en los programas del horario de mayor audiencia.*

Varias características hacen que algunas hipótesis sean mejores que otras, y hablaremos de ellas en el siguiente capítulo.

Por ahora, debemos tener presente que una hipótesis es una extensión objetiva de la pregunta que se planteó originalmente. Si bien es posible que no todas las preguntas tengan una respuesta por la forma en que se plantean, lo cual está bien en la etapa de la pregunta, una buena hipótesis plantea una pregunta en una forma *que puede probarse*. Las buenas preguntas llevan a buenas hipótesis, que a su vez dan pie a buenos estudios.

Recopilación de la información pertinente

Las hipótesis deben plantear una relación clara entre diferentes factores como ver televisión y un comportamiento agresivo en los hombres. Ésa es la tarea de la hipótesis. Una vez formulada una hipótesis, el siguiente paso es obtener información o datos empíricos que confirmen o refuten la hipótesis. Si estamos interesados en saber si ver programas violentos de televisión causa o no causa un comportamiento agresivo, es necesario recabar datos de tal índole que permitan probar la hipótesis.

Por ejemplo, podríamos recopilar dos tipos de datos para probar la hipótesis antes mencionada. El primero podría ser el número de actos violentos en un segmento de una hora del horario televisivo de máxima audiencia. El segundo podría ser el número de comportamientos agresivos observados en niños que miraron el programa y el número de tales actos en niños que no miraron el programa.

Un punto importante respecto a someter a prueba las hipótesis es que un buen científico busca *probarlas*, no necesariamente *demostrarlas*. Como buen científico, usted debe concentrarse en recabar datos que revelen tanto acerca de la verdad del mundo como sea posible, permitiendo que acontezcan los hechos, sea que la hipótesis resulte "correcta" o "equivocada", y esté usted de acuerdo o no con los resultados. Proponerse demostrar una hipótesis puede hacer que los científicos predispongan los métodos de recopilación de datos o las formas de interpretar los resultados. Si hay predisposición, toda la secuencia de pasos puede venirse abajo. Además, en ciencia no es realmente posible estar "equivocado". No encontrar apoyo para una hipótesis sólo significa que hay más preguntas que hacer o que es preciso reformular las preguntas que se hicieron. Esa es la belleza de la buena ciencia: siempre hay otra pregunta que hacer sobre el mismo tema y que puede arrojar un poco más de luz sobre el asunto. Y, ¿quién sabe? Ese poquito más de luz podría ser justo lo que se necesita para descubrir un hallazgo importante y totalmente nuevo.

Prueba de la hipótesis

¿Basta simplemente con recabar datos relacionados con los fenómenos que se estudian? De ninguna manera. ¿Qué tal si usted ya terminó de recopilar datos y encuentra que los muchachos que vieron programas de televisión agresivos durante el horario de máxima audiencia exhiben 4.8 actos agresivos en el periodo de una hora después de la exposición, y que los muchachos que miraron un programa no agresivo exhibieron en promedio 2.2 actos? ¿Cuál

sería su conclusión?

Por un lado, usted podría decir que los muchachos que miraron los programas agresivos mostraron más del doble de agresividad. Por otra parte, podríamos argüir que la diferencia entre los dos promedios no es lo bastante grande como para llegar a una conclusión. Para afirmar que ver el segmento de televisión agresivo fue en verdad determinante, tendríamos que observar una diferencia mucho mayor. ¿Es esto un dilema insoluble? De ninguna manera.

Aquí es donde entra en acción la *estadística inferencial* (que estudiaremos más a fondo en el capítulo 8), un conjunto de herramientas que permiten a los investigadores separar los efectos de un factor aislado (como ver televisión agresiva o no agresiva) de las diferencias entre grupos que podrían deberse a algún otro factor o sencillamente al **azar**. Sí, la suerte, la fatalidad, el destino, la fortuna o como sea que usted quiera llamar a lo que no se puede controlar, a veces puede ser la causa de diferencias entre grupos. Por ejemplo, ¿qué tal si uno de los muchachos que no miró el segmento agresivo anda de malas ese día y decide darle un golpe a su amiguito? ¿O si uno de los muchachos que miró el segmento agresivo está cansado y simplemente no tiene ganas de jugar? La tarea de estas herramientas es ayudarnos a separar los efectos de los factores que estamos estudiando de otros factores no relacionados. Lo que hacen estas herramientas estadísticas es asignar un nivel de probabilidad a un resultado para poder decidir si lo que vemos en verdad tiene la causa que creemos o se debe a alguna otra cosa que habrá que dejar para el siguiente estudio.

Trabajo con la hipótesis

Una vez que usted ha recopilado los datos que su pregunta requiere y que ha probado la hipótesis, como buen científico se sienta, se pone cómodo, adopta un aire de intelectual, y examina los resultados. Éstos podrían confirmar o refutar la hipótesis. En cualquier caso, el camino está claro. Si se obtuvo confirmación, ya se sabe que, como se había sugerido, los factores tienen una relación entre sí y son conceptualmente importantes, y usted puede dedicarse alegremente a planear su siguiente experimento científico. Si la hipótesis no se confirmó, bien puede ser una oportunidad para aprender algo que antes no se sabía. En el ejemplo que usamos antes, esto significaría que mirar segmentos televisivos con modelos agresivos por sí solo no causa un comportamiento agresivo por parte de los muchachos. Aunque el investigador tal vez se sienta un poco decepcionado por no haber podido apoyar su presentimiento inicial (llamado formalmente hipótesis), los resultados de un estudio bien realizado siempre proporcionan información valiosa, sea cual sea el resultado.

Reconsideración de la teoría

Por último, es hora de hacer un balance y relacionar todas estas labores de investigación con lo que guía fundamentalmente dichas labores: la teoría. Ya definimos una teoría como una serie de enunciados que predicen cosas que ocurrirán en el futuro y explican cosas que han ocurrido en el pasado. Pero la naturaleza misma de las teorías es que pueden modificarse según los resultados de investigaciones basadas en los mismos supuestos en los que se basa la teoría.

Por ejemplo, un enfoque específico para entender el desarrollo de niños y adultos se denomina teoría del aprendizaje social, que da especial importancia al papel del modelamiento y del aprendizaje vicario o indirecto. Según esta teoría, la exposición a un comportamiento agresivo daría pie a un comportamiento agresivo, una vez que el entorno contiene los mismos tipos de indicios que estaban presentes cuando se observó el modelo agresivo inicial (como personajes agresivos de dibujos animados).

Si se confirma la hipótesis de que observar tales modelos aumenta la agresión, habremos añadido otro bloque de construcción, otra prueba, a la casa que llamamos teoría del aprendizaje social. Los buenos científicos siempre están tratando de ver qué tipo de tabique (nueva información) cabe en qué lugar, o si no cabe. De este modo, los conocimientos nuevos pueden modificar el aspecto de la teoría (o la casa) o lo que tiene que decir acerca del comportamiento humano. Así, podrían generarse, a partir de la teoría, nuevas preguntas que contribuyan también a la estructura de la casa.

Hacer nuevas preguntas

En todo caso, el último paso de este modelo simplificado de investigación científica es hacer una nueva pregunta. Ésta podría ser una sencilla variación sobre el tema (*¿Los hombres reaccionan de forma diferente que las mujeres ante modelos agresivos?*) o una refinación de la pregunta original (*¿Qué tanta exposición a los modelos agresivos es necesaria para que los niños comiencen a modelar el comportamiento?*) No importa si la hipótesis se apoya o se

rechaza; la buena investigación nos deja más adelantados en el camino hacia la respuesta a la pregunta original. Lo único es que tal vez estemos en un lugar diferente de aquel en el que pensamos que estaríamos.

Diferentes tipos de investigación

A estas alturas, ya usted tiene una buena idea de qué es la investigación y cómo funciona el proceso de investigar. Es hora de poner nuestra atención en una descripción y en ejemplos de diferentes tipos de métodos de investigación y de la clase de preguntas que plantean.

Es interesante que los tipos de métodos de investigación que estudiaremos difieran principalmente en dos dimensiones: la naturaleza de la pregunta que se hace y el método empleado para contestarla. Pero en algo en lo que estos métodos no necesariamente difieren es en el contenido o enfoque de la investigación. Dicho de otro modo, si lo que a usted le interesa son los efectos de la televisión en los niños, su investigación puede ser no experimental, reseñando los hábitos de ver televisión, o experimental, exponiendo niños a ciertos modelos y observando el efecto de la exposición sobre su comportamiento.

En la tabla 1.1 se muestra un resumen de las tres principales categorías de métodos de investigación que se estudiarán en este libro. En la tabla puede verse el propósito de cada una, el marco referencial de tiempo en el que se desarrolla cada una, el grado de control que los diferentes métodos tienen sobre los factores en competencia, y un ejemplo de cada categoría. En los capítulos 9, 10 y 11 entraremos en más detalles de cada uno de estos métodos de investigación.

Investigación no experimental

La investigación no experimental incluye diversos métodos que describen relaciones entre variables. La distinción importante entre los métodos no experimentales y los demás que mencionaremos posteriormente es que los métodos de investigación no experimentales no establecen, ni pueden probar, relaciones causales entre variables. Por ejemplo, si quisiéramos reseñar el comportamiento de ver televisión de los adolescentes, podríamos hacerlo pidiéndoles mantener un diario en el que anotaran lo que ven y con quién lo ven. Semejante estudio descriptivo proporcionaría información acerca de sus hábitos de ver televisión pero nada dice acerca de por qué ven lo que ven. No estamos tratando de influir de manera alguna sobre su comportamiento de ver televisión ni investigar por qué podrían ver ciertos programas. La naturaleza de esta investigación es no experimental porque no se están haciendo hipótesis respecto a relaciones de causa y efecto de ningún tipo.

Los métodos de investigación no experimental que se cubrirán en este libro son el descriptivo, el histórico y el correlacional. Los tres se tratarán en el capítulo 9. A continuación presentamos una breve descripción global de cada uno.

Tabla 1.1 Los diferentes tipos de investigación que se estudiarán en este libro.

Tipo de investigación	No experimental (descriptiva)	No experimental (histórica)	No experimental (correlacional)	Experimental	Cuasiexperimental
Propósito	Describir las características de un fenómeno existente	Relacionar sucesos que han ocurrido en el pasado	Examinar las relaciones entre variables	Probar si hay verdaderas relaciones de causa y efecto	Probar si hay relaciones causales, sin tener pleno control
Marco referencial de tiempo	Actual	Pasado	Actual o pasado (correlación)	Actual	Actual o pasado
Grado de control sobre los factores	Ninguno o bajo	Ninguno o bajo	Bajo o mediano	Alto	Moderado o alto
Ejemplo	Encuesta de prácticas de cortejo de mujeres adolescentes	Análisis del uso de la hipnosis, por Freud y su relación con las prácticas psicoterapéuticas actuales escolares.	Investigación de la relación entre el número de horas de ver TV y el promedio de calificaciones	El efecto de un programa de lenguaje preescolar sobre las capacidades de barrios céntricos de las ciudades	Diferencias de género en las habilidades espaciales

Investigación descriptiva

La investigación descriptiva reseña las características de un fenómeno existente. Los censos nacionales son investigaciones descriptivas, lo mismo que cualquier encuesta que evalúe la situación actual de cualquier aspecto, desde el número de grifos en las casas hasta el número de adultos de más de 60 años que tienen nietos.

¿Qué puede hacerse con esta información? Primero, obtener una imagen amplia de un fenómeno que nos podría interesar explorar. Por ejemplo, si a usted le interesa aprender más acerca del proceso de lectura en los niños, podría querer consultar documentos publicados por las autoridades de educación u organismos del ramo. Por ejemplo, podría existir un documento que resume el aprovechamiento en cuanto a lectura de niños de diferentes grupos de edades. O bien, podrían consultarse publicaciones de las autoridades de salubridad para averiguar la proporción de casos de paperas en alguna región del país, o las estadísticas de las autoridades laborales para conocer la tasa de desempleo actual y cuántos padres o madres que trabajan y no tienen cónyuge tienen hijos de menos de 5 años de edad (cerca de 60%). Si usted quiere saber algo, hay un lugar donde puede averiguarlo. La investigación descriptiva exige este tipo de información.

Otro ejemplo son las entrevistas que Peter O. Peretti y Kris G. Majecen (1991) hicieron a 58 individuos de edad avanzada, entre 68 y 87 años, empleando un esquema estructurado para investigar las variables que afectan el abuso emocional entre los ancianos. Como resultado de las entrevistas (un tipo de investigación de reseña), encontraron nueve variables comunes al abuso de ancianos, que incluyen falta de afecto, amenazas de violencia y confinación.

La investigación descriptiva no sólo puede ser autosuficiente, como demuestran los ejemplos, sino también puede servir como base para otros tipos de investigaciones, porque a menudo es preciso describir las características de un grupo antes de poder abordar la significatividad de cualesquier diferencias observadas.

Investigación histórica

La **investigación histórica** relaciona sucesos del pasado con otros acontecimientos de la época o con sucesos actuales. Básicamente, la información histórica (o **historiografía**) contesta la pregunta: *¿Cuál es la naturaleza de los acontecimientos que han ocurrido en el pasado?* Por ejemplo, tal vez quisiéramos examinar tendencias en la forma en que se han tratado las enfermedades mentales o cómo han cambiado las actitudes hacia el trabajo y la familia. Todo esto requiere el trabajo de detective de un historiador para hallar y recopilar datos pertinentes y luego, al igual que en cualquier otro empeño de investigación, probar una hipótesis. De hecho, lo mismo que cualquier otro investigador, el historiador recaba datos, los analiza y llega a conclusiones acerca de qué tan defendible es su hipótesis. Una diferencia significativa entre la investigación histórica y otros tipos de investigaciones está en el *tipo* de datos recabados y el método para recopilarlos.

Quienes se dedican a la investigación histórica a menudo logran sus fines empleando *fuentes primarias* (documentos originales o personas que han experimentado personalmente un suceso) y *fuentes secundarias* (documentos de segunda mano o personas que podrían tener cierto conocimiento acerca del suceso pero que no estaban presentes cuando ocurrió). No obstante, aunque tales fuentes estén fácilmente accesibles, uno de los mayores retos para la realización de tales investigaciones está en saber qué tanto el investigador puede confiar en la veracidad y exactitud de las fuentes.

Un ejemplo de investigación histórica es el examen que hicieron Nancy Burton y Lyle Jones (1982) de las tendencias en los niveles de aprovechamiento de niños de raza negra y blanca. Burton y Jones examinaron las proporciones de graduados de bachillerato para negros y blancos nacidos antes de 1913, entre 1913 y 1922, 1923 y 1932, 1933 y 1942, 1943 y 1947, y 1948 y 1952, para personas que tenían 25 años o más en 1977. También examinaron varios otros indicadores históricos en grupos más recientes de niños blancos y negros y concluyeron que las diferencias en el aprovechamiento entre estudiantes blancos y negros están disminuyendo. Para completar su análisis, Burton y Jones obtuvieron datos de la Evaluación Nacional del Progreso Educativo de Estados Unidos. Con las avanzadas herramientas de recuperación de datos de que se dispone hoy día, los investigadores históricos pueden usar sus computadoras (véase el capítulo 3) para consultar casi cualquier base de datos que necesiten y ahorrarse visitas a las bibliotecas.

Investigación correlacional

Las investigaciones descriptiva e histórica proporcionan una imagen de los sucesos que están ocurriendo o que han ocurrido en el pasado. En muchos casos los investigadores desean ir más allá de la mera descripción para analizar la relación que podría existir entre ciertos sucesos. El tipo de investigación que con mayor probabilidad podría responder a preguntas acerca de la relación entre variables o sucesos se llama investigación correlacional.

Una diferencia de la **investigación correlacional** respecto de las investigaciones descriptiva e histórica, es que proporciona indicios de la relación que podría existir entre dos o más cosas, o de qué tan bien uno o más datos podrían predecir un resultado específico. La investigación correlacional utiliza un índice numérico llamado coeficiente de correlación (que trataremos con detalle en el capítulo 9) como medida de la fortaleza de tal relación. En casi todos los estudios correlacionales se informa el valor de dicho índice.

Si a usted le interesara encontrar la *relación* entre el número de horas que los estudiantes de primer año de licenciatura estudian y su promedio de calificaciones, estaría realizando una investigación correlacional porque lo que le interesa es la relación entre estos dos factores. Si quiere averiguar cuál es el mejor conjunto de *predictores* del éxito en las escuelas de posgrado, estaría realizando un tipo de investigación correlacional que incluye predicción.

Por ejemplo, en un estudio de la relación entre temperamento y conductas de apego en bebés (Vaughn, Lefever, Seifer y Barglow, 1989), los investigadores examinaron la correlación entre diferentes tipos de conductas de apego (qué tan vinculados están los bebés a su madre) y el temperamento general del bebé, un término que suele utilizarse para analizar la personalidad del ser humano en su etapa temprana. Los investigadores encontraron que el temperamento de un bebé no predice qué tan apegado está el niño a su madre.

Uno de los puntos más importantes respecto a la investigación correlacional es que examina relaciones entre variables pero de ningún modo implica que una es la causa de la otra. En otras palabras, la correlación y la predicción examinan asociaciones pero no relaciones causales, donde un cambio en un factor influye directamente en un cambio en otro. Por ejemplo, es un hecho bien establecido que a medida que aumenta el número de delitos en una comunidad, sucede lo mismo con el nivel de consumo de helados! ¿Qué está sucediendo? Sin duda, ninguna persona que razone normalmente llegaría a la conclusión de que existe una relación causal entre ambas cosas, de modo que si se prohibieran los helados desaparecerían los delincuentes. Más bien, otra variable, la temperatura, explica mejor la cantidad de helado consumido y la tasa delictuosa (ambas aumentan cuando hace calor). Podría parecer ridículo que la gente fuera a identificar una causalidad sólo porque existe una relación entre sucesos, pero no hay que adentrarse mucho en la lectura del periódico matutino para ver cómo muchos políticos llegan a tales conclusiones imprudentes.

Investigación experimental

Ya sabemos que la investigación correlacional puede ayudar a establecer la presencia de una relación entre variables pero sin darnos alguna razón para creer que existe una relación causal entre ellas. ¿Cómo averiguamos si ciertas características, comportamientos o sucesos están relacionados de tal manera que la relación es causal? Hay dos tipos de investigación que pueden contestar esa pregunta. La primera es la investigación cuasiexperimental y la segunda es la investigación experimental. Por ahora, examinaremos brevemente esta última.

La única forma de establecer una verdadera relación de causa y efecto en cualquier estudio es aislar y eliminar todos los factores que podrían ser la causa de un resultado en particular y probar tan sólo los que se quiere medir directamente.

La investigación experimental es aquella en la que los participantes se asignan a grupos con base en algún criterio determinado que suele llamarse variable de tratamiento. Por ejemplo, supongamos que a usted le interesa comparar los efectos de dos técnicas distintas para reducir el comportamiento de desorden obsesivo compulsivo en adultos. La primera técnica incluye terapia del comportamiento; la segunda no. Una vez que se han asignado los adultos a grupos y que han concluido los programas, usted querrá buscar cualesquier diferencias entre los dos grupos en cuanto a los efectos de la terapia sobre el número de comportamientos obsesivos compulsivos. Puesto que quien determina la asignación a grupos es el investigador, él o ella controla totalmente a qué se exponen los adultos.

Este es el entorno ideal para establecer una relación de causa y efecto, porque se ha definido con claridad la posible causa (si en realidad produce algún efecto) y se puede vigilar de cerca lo que está sucediendo. Pero lo más importante es que se tiene control total sobre el tratamiento. En un estudio cuasiexperimental, no se tiene un grado tan alto de control, porque la gente ya se ha asignado indirectamente a los grupos (por clase social, abuso, género y tipo de daño) para los cuales se están probando los efectos.

La distinción entre los métodos de investigación experimentales y de otro tipo se reduce a la cuestión del control. Los verdaderos diseños de investigación experimental, que estudiaremos en el capítulo 10, aíslan y controlan todos los factores que podrían causar algún efecto, con excepción de aquel que más interesa.

Por ejemplo, Fleming, Klein y Corter (1992) examinaron los efectos de participar en un grupo de apoyo social contra la depresión, las actitudes maternas y el comportamiento de las nuevas madres. Como parte del diseño experimental, los investigadores dividieron un grupo de 142 madres en tres subgrupos. El grupo 1 recibió la intervención, el grupo 2 recibió la condición de no intervención y el grupo 3 recibió una intervención especial de "grupo por correspondencia". El punto clave aquí es la *manipulación* (la palabra clave en los diseños experimentales) de la condición para cada uno de los tres grupos.

Esta investigación es experimental, ya que el investigador determinó la participación de los miembros de los grupos en el grupo de apoyo social en función del tratamiento mismo. Como veremos, en un estudio cuasiexperimental el investigador no controla quién entra en cuál grupo.

La diferencia primaria entre la investigación cuasiexperimental y la experimental es que en **la investigación cuasiexperimental** el investigador no tiene control total sobre el criterio empleado para asignar participantes a grupos, pero en la investigación experimental sí lo tiene. La diferencia es enorme.

En la investigación cuasiexperimental los participantes se asignan a grupos con base en alguna característica o cualidad que estas personas aportan al estudio. Ejemplos de ello son las diferencias de sexo, edad, grado escolar, vecindario, tipo de trabajo e incluso experiencias. Estas asignaciones a grupos ocurren *antes* de iniciarse el experimento, y el investigador no puede controlar quién pertenece a cada grupo.

Digamos que a usted le interesa examinar los patrones de voto en función del vecindario. No es posible cambiar el vecindario en el que la gente vive, pero sí podemos usar el método cuasiexperimental para establecer una relación causal entre el lugar de residencia y los patrones de voto. Dicho de otro modo, si averiguamos que el patrón de voto y el lugar de residencia están relacionados, podemos decir con cierto grado de confianza (aunque no tanta como en un estudio experimental) que el lugar en que alguien reside tiene cierta relación causal con la forma en que esa persona vota.

El uso más importante del método cuasiexperimental es en los casos en que los investigadores no pueden, con la conciencia tranquila, asignar a la gente a grupos y probar los efectos de la pertenencia a grupos sobre algún otro resultado. Por ejemplo, los investigadores interesados en los efectos del desempleo sobre los niños no podrían alentar a sus madres o padres a que renunciaran a su trabajo. Más bien, buscarían familias en las que los padres ya están desempleados y luego realizarían la investigación. Norma Radin y Rena Harold-Goldsmith (1989) hicieron exactamente eso: compararon la relación entre padres e hijos para 17 padres sin empleo y 31 padres con empleo. Los investigadores examinaron además otros factores, como la opinión que el padre tiene del papel masculino en la familia, el hecho de que la madre trabaje, y la edad del niño.

La investigación cuasiexperimental también se denomina investigación **post hoc** o investigación después del hecho, porque la investigación misma se efectúa después de la asignación a grupos (como empleado o desempleado, desnutrido o con nutrición normal, hombre o mujer). Puesto que la asignación ya se ha realizado, el investigador tiene un alto grado, pero no el grado máximo, de control sobre la causa de cualesquier efectos que se estén examinando. Para tener el grado más alto de control, es preciso usar el método experimental.

Investigación básica versus aplicada

En el mundo de la investigación a veces es necesario hacer distinciones no sólo acerca del tipo de investigación sino también acerca de la categoría más general a la que podrían pertenecer las implicaciones o la utilidad de la investigación. Es aquí donde entra la distinción entre investigación básica y aplicada. Pero, ¡cuidado! Hay ocasiones en que se utiliza esta distinción como forma cómoda de clasificar las actividades de investigación y no para iluminar la intención o propósito del investigador y la importancia del estudio.

La distinción más fundamental entre las dos es que la **investigación básica** (también llamada investigación pura) es investigación que no tiene una aplicación inmediata en el momento en que se termina, mientras que la **investigación aplicada** sí la tiene. Si ésta le parece una distinción un tanto ambigua, es porque lo es, ya que casi todas las investigaciones básicas tarde o temprano conducen a alguna aplicación valiosa a largo plazo.

Por ejemplo, cada dólar gastado en la investigación básica que apoyó las misiones lunares durante las décadas de los sesenta y los setenta redituó seis dólares en impacto económico. Datos de investigaciones básicas que postulan una relación entre el mal de Alzheimer en personas de edad avanzada y el síndrome de Down (un desorden genético) en personas más jóvenes podría, con el tiempo, ser el hallazgo crítico que lleve a una cura para ambas enfermedades. Otro ejemplo: A quién le importa si a algunos niños les cuesta más trabajo que a otros distinguir entre dos estímulos muy similares? A usted, si es que quiere enseñarles a leer. Muchos programas de lectura se han desarrollado directamente a partir de labores de investigación básica como ésta.

Por tanto, no debemos juzgar ni la calidad del producto terminado ni el mérito de apoyar un proyecto de investigación rotulándolo como investigación básica o aplicada. Más bien, hay que examinar detenidamente su contenido, y juzgarlo por su valor intrínseco. Es evidente que hay quienes siguen este consejo, pues cada vez aparecen más informes sobre investigación básica (que en alguna época se consideró más allá del interés de los practicantes ordinarios) en publicaciones profesionales orientadas hacia dichos practicantes, así como en semanarios de noticias y en revistas de ciencia popular.

¡Magnífico! Ya ha terminado usted el primer capítulo y cabe esperar que se haya formado un concepto claro acerca de qué es (y qué no es) la investigación, cuáles son sus fines y de qué maneras se puede realizar. Armado con esta nueva información, pase ahora al siguiente capítulo, que se concentra en la jerga que los investigadores usan y en la forma como esos nuevos términos cuadran con lo que usted ha aprendido aquí.

Ejercicios

1. El proceso de investigación nunca es independiente del contenido del tema de investigación. Como estudiante que se inicia en el campo de la investigación, y tal vez incluso en su propia disciplina (como psicología, sociología o enfermería), conteste las preguntas siguientes:
 - a. ¿Qué áreas dentro de su disciplina tienen un interés especial para usted?
 - b. Mencione algunos investigadores sobresalientes en su campo e indique el enfoque de sus trabajos.
 - c. De los diferentes tipos de investigación que describimos y analizamos en el capítulo, ¿cuál cree usted que mejor se ajuste al tipo de investigaciones que se realizan en su disciplina?
2. Visite la biblioteca de su colegio o universidad y encuentre un artículo de una revista profesional que describa un estudio de investigación. Con base en la descripción del proceso de investigación científica (que presentamos en este capítulo), conteste las preguntas siguientes:
 - a. ¿Cuál es la pregunta primaria que plantea el estudio?
 - b. ¿Qué factores importantes se identifican?
 - c. ¿Se plantea una hipótesis? Si se hace, ¿cuál es?
 - d. Describa la forma en que se recabó la información.
 - e. ¿Cómo podrían los resultados del estudio afectar la hipótesis que se planteó originalmente?
3. Entreviste a un investigador activo de la universidad de su comunidad y pregúntele acerca de sus actividades de investigación, incluyendo
 - a. el enfoque de sus intereses en investigación,
 - b. la razón por la que le interesa esa área,
 - c. qué parte de la investigación es la más estimulante,
 - d. qué parte de la información es la menos estimulante, y
 - e. qué impacto podrían tener sobre su disciplina específica los resultados de la investigación.
4. Escoja una disciplina dentro de las ciencias sociales *y del comportamiento, como desarrollo infantil, psicología infantil, educación superior o psicología de salud*; Dentro de la disciplina seleccionada, encuentre un estudio representativo de naturaleza cuasiexperimental o experimental, redacte una descripción del estudio que ocupe un párrafo. Haga lo mismo con un estudio histórico.
5. En un estudio correlacional ficticio, los resultados mostraron que la edad esta relacionada con la fuerza. Esto es, a medida que los niños crecen, su fuerza aumenta. ¿(defecto tiene la afirmación de que *el aumento en la fuerza es causado por el aume en la edad* o que *cuanto más fuerte se vuelve una persona, más vieja se hace*?

6. Para usted, ¿cuál sería una buena definición de ciencia? ¿En qué diferiría su definición de la ciencia respecto a la de un estudiante de una clase similar hace 25 años? ¿En qué diferiría su definición de la ciencia dada por un científico físico (por ejemplo, un físico, un químico), si acaso difiere?

7. Busque ejemplos de editoriales o artículos de información que presenten datos correlacionales. ¿Los autores infieren una relación de causa y efecto en la correlación? ¿qué podría ser difícil incluso para investigadores experimentados abstenerse de cometer tal error?

8. La investigación a menudo repite los hallazgos hechos por otros. ¿Qué valor tiene el proceso?

9. Identifique cinco atributos que caracterizan la investigación de alta calidad.

10. Expresar la pregunta siguiente como hipótesis:

¿Qué impacto social tienen las guarderías infantiles sobre un niño que asiste a ella durante sus años de educación preescolar y primaria?

11. Explique la diferencia entre investigación histórica, correlacional y cuasiexperimental.

CAPITULO DOS **El Proceso de Investigación**

En Neil J. Salkind “Métodos de Investigación”, Prentice Hall, México, 1998.

Lo que aprenderá en este capítulo

- El camino desde la formulación de preguntas hasta la búsqueda y el hallazgo de soluciones
- La diferencia entre variables dependientes e independientes
- Qué es una hipótesis, y cómo funciona
- La importancia de la hipótesis nula
- La diferencia entre la hipótesis nula y la de investigación
- Las características de una buena hipótesis
- La importancia de las muestras y poblaciones en el proceso de investigación
- Qué es la significancia estadística
- Qué criterios debemos usar para evaluar un artículo de investigación
- Cómo las prácticas de investigación éticas son responsabilidad de todos los investigadores
- Por qué la confidencialidad es uno de los criterios éticos que es más importante mantener
- Las normas éticas publicadas por diferentes organizaciones profesionales
- La importancia de los comités de revisión de prácticas de investigación internas

Del problema a la solución

Todo lo que necesitamos es una pregunta interesante, recabar algunos datos y ¡listo! ¡Investigación instantánea! No exactamente. El modelo de investigación científica que estudiamos en el capítulo 1 especifica muy bien los pasos del proceso de investigación, pero éste tiene varios aspectos adicionales.

Al principio de este capítulo presentaremos un ejemplo "de la vida real" de cómo se lleva a cabo realmente el proceso y cómo los investigadores comienzan con lo que consideran un problema (por resolver) y terminan con una solución (o los resultados) para ese problema, hay que tener presente que el significado de las palabras *problema* y *solución* va más allá de la respuesta a un problema sencillo del tipo $2 + 2$. Más bien, las preguntas que los investigadores hacen a menudo reflejan una preocupación social urgente o una cuestión económica. Además, los resultados de un estudio de investigación como éste proporcionan las bases para la siguiente empresa de investigación.

Examinaremos un interesante estudio intitulado *Empleo materno y experiencias diarias de adolescentes jóvenes en*

familias que viven sólo con la madre (Duckett y Richards, 1989) que examina el impacto del empleo materno sobre el desarrollo de los adolescentes. Uno de los aspectos más creativos de este estudio es la forma en que los investigadores recopilaban los datos que necesitaban. No se sentaron con los adolescentes y les preguntaron sus opiniones acerca de esto o lo otro, sino más bien trataron de obtener una imagen general de sus sentimientos fuera del entorno de laboratorio.

Los investigadores que realizaron este estudio con 436 jóvenes de quinto año de primaria hasta tercero de secundaria, y sus madres, estaban interesados en una combinación de aspectos que han recibido considerable atención en los medios impresos y electrónicos. El objetivo general de la investigación (y el problema) era entender mejor algunos de los factores y consecuencias que rodean al creciente número de madres que trabajan y tienen hijos adolescentes.

Para enfocar su investigación, se propusieron determinar la naturaleza general de las experiencias de los jóvenes en función de tener una madre que trabaja, así como la calidad del tiempo que los adolescentes pasaban con sus madres. En vista de que tantas madres, tanto de familias con ambos padres como de familias con uno solo, están trabajando ahora fuera del hogar, las respuestas a preguntas como las planteadas por este estudio están adquiriendo cada vez más importancia en la formación de políticas económicas y sociales.

Para obtener las respuestas que querían, los investigadores tuvieron que comparar adolescentes que vivían con ambos padres (382, o sea 88% del total) con adolescentes que vivían únicamente con su madre (54, es decir, 12% del total). Pero para lograr plenamente su objetivo de entender mejor los efectos del empleo materno, los investigadores tuvieron que subdividir el grupo de jóvenes y padres, en aquellos jóvenes cuyas madres trabajaban medio tiempo, tiempo completo o estaban desempleadas. Al separar los sujetos en grupos con base en estos factores (configuración familiar y situación de empleo), los investigadores pudieron hacer comparaciones dentro y entre los seis grupos (todas las combinaciones de familias con uno y dos padres, con madres empleadas medio tiempo, tiempo completo o desempleadas) y obtener la información que necesitaban para responder las preguntas generales arriba planteadas.

Ahora viene la parte verdaderamente creativa del estudio. Los investigadores usaron un método llamado *método de muestreo por experiencia* desarrollado inicialmente por otros dos investigadores (Csikszentmihalyi y Larson, 1987). Según este método, los adolescentes que participan en el estudio llevan notificadoros (*beepers*) electrónicos. A intervalos impredecibles, los adolescentes reciben un aviso de la central y en ese momento deben dejar de hacer lo que están haciendo y llenar un formato de informe sobre ellos mismos. Esto se realizó durante una semana.

La señal que indicaba al participante detenerse y llenar el formato se envió en promedio cada dos horas entre las 7:30 a.m. y las 9:30 p.m., enviándose un total de 49 señales durante la semana a cada participante. Esto implica que en el curso de una semana se llenaron 49 formatos individuales, lo que proporcionó abundante información sobre lo que lo participantes estaban sintiendo en un momento dado. Con 436 participantes y 49 formatos por participante, se obtuvo un total de 21 364 formatos, lo cual es una muestra sustancial del comportamiento de jóvenes.

¿Qué contenían esos formatos de requisición de información sobre los individuos que 105 contestaban? Los adolescentes tenían que informar acerca de lo que los investigadores llaman *afecto* (contento-triste, animoso-irritable, amigable-enojado) y *estimulación* (alerta-somnoliento, fuerte-débil, excitado-aburrido). Cada uno de estos seis elementos se calificó empleando una escala del 1 al 7. Por ejemplo, un participante podría anotar un 4 para indicar que se sentía "justo a la mitad entre contento y triste en ese momento". Estos seis elementos pueden calificarse en muy poco tiempo, y así se obtuvo una imagen exacta de la vida diaria del adolescente. Los jóvenes también tenían que contestar a las preguntas "¿Qué estabas haciendo?" y "¿Con quién estabas", y a otras acerca de su percepción de qué tan amigables estaban sus padres y qué sentían ellos mientras estaban con sus padres.

Duckett y Richards tenían entonces una comparación interesante (madres con y sin cónyuge que están desempleadas o empleadas medio tiempo o tiempo completo) y una buena cantidad de reacciones de los adolescentes, sobre lo cual basar su análisis y discusión. Para entender esta información, los investigadores la compilaban y luego le aplicaron ciertas pruebas estadísticas (ya aprenderá más acerca de éstas más adelante) para sacar sus conclusiones. Entre estas últimas estuvieron

- Los hijos de madres solas que trabajan se benefician en otras formas aparte de la obtención de ingresos
- El empleo materno está relacionado con interacciones madre-hijo positivas.
- los hijos de madres solas con empleo de tiempo completo sienten más amistad hacia sus papás.

El anterior fue un buen estudio que examinó una pregunta relacionada con muchas cuestiones para las que todo mundo, desde los maestros de escuela hasta las empresas, necesita respuestas. El estudio se realizó con un grupo de

participantes de tamaño más que suficiente y utilizó métodos que obtuvieron en forma directa el tipo de información que los investigadores querían. Aunque no se obtuvieron respuestas a todas las preguntas acerca de la relación entre el empleo de la madre y el desarrollo de los adolescentes, los investigadores aportaron una pieza importante al rompecabezas que implica entender los efectos del empleo sobre los niños en crecimiento y las familias cambiantes. Los investigadores evidentemente adoptaron un enfoque lógico para partir desde una pregunta que tiene importancia para muchos grupos de la sociedad actual, articulándola de tal manera que se pudo contestar en una forma razonable y eficiente. Sin duda, la cuestión de cómo afecta a los niños el empleo de sus padres persiste, pero los resultados de investigaciones como la que hemos resumido nos acercan a una solución para algunos de los problemas planteados por tales situaciones de trabajo. Para ser el tipo de investigador que usted desea ser, necesita conocer las reglas del juego y seguirlas como hicieron Duckett y Richards. Estos conocimientos parten de la comprensión de cierta terminología e ideas básicas.

El lenguaje de la Investigación

¿Niveles de significancia? ¿Hipótesis nula? ¿Variables independientes? ¿Diseños factoriales? ¿Hipótesis de investigación? ¿Muestras? ¿Poblaciones?

Éstas y otras palabras y términos nuevos constituyen la base de gran parte de la comunicación que tiene lugar en el mundo de la investigación. Al igual que en cualquier ambiente, es difícil participar si no se conocen las reglas. Las reglas comienzan aquí, con una comprensión básica de la terminología que los investigadores utilizan en sus actividades cotidianas. Este capítulo es una especie de lección de idiomas. Una vez que usted se familiarice con estos términos, el resto del libro será más fácil de entender y más útil. Todos los términos que describiremos y definiremos aquí se usarán una y otra vez a lo largo de esta obra.

Todo acerca de las variables

La **palabra variable** tiene varios sinónimos, como cambiante o inestable. nuestro conjunto de reglas nos dice que una variable es un sustantivo, no un adjetivo, y representa una clase de resultados que pueden asumir más de un valor.

Por ejemplo, el color del pelo es una variable que puede adoptar los valores de rojo, castaño, negro, rubio y, en nuestros días, verde, anaranjado y (en serio) morado. (¡Lo vi afuera de mi oficina esta mañana!). Otras variables serían altura (alto o bajo), peso (60 kilos o 70 kilos), edad de vacunación (6 semanas o 18 meses), número de palabras recordadas, ausencia del trabajo, afiliación a un partido, etcétera. Lo único que estos rasgos tienen en común es que la variable (como la afiliación a un partido) puede tomar cualquiera de varios valores (conservador, liberal, socialdemócrata, comunista, etcétera).

Resulta interesante que variables que podrían tener el mismo nombre no sólo pueden asumir valores distintos podríamos medir la estatura en centímetros (170 cm) o por rango (el más alto), por ejemplo-, sino también definirse de diferente manera, dependiendo de muchos factores, como el propósito de la investigación o las características de los participantes. Por ejemplo, consideremos la variable llamada inteligencia. Para un investigador la definición podría ser las calificaciones de la Prueba de Inteligencia Stanford-Binet, mientras que para otro podría ser el puntaje obtenido en la Batería de Evaluaciones Kaufmann. Para alguien como Howard Gardner (1983), quien cree en la existencia de inteligencias múltiples, la definición podría ser el desempeño en matemáticas, música o en alguna actividad física. Todas estas variables representan el mismo constructo general de inteligencia, evaluado de diferentes maneras.

En los próximos párrafos describiremos varios tipos de variables. En la tabla 2.1 se resumen dichos tipos y sus características.

Tipo de variables	Definición	Sinónimos
Dependiente	Variable que indica si el tratamiento o manipulación de la variable independiente tuvo algún efecto.	Variable resultante Variable resultante Efecto Variable de criterio
independiente	Variable que se manipula para examinar su impacto en una variable dependiente	Tratamiento Factor Variable predictora

o resultante.

De control	Variable relacionada con la variable dependiente y cuya influencia es preciso eliminar.	Variable restrictora
Extraña	Variable relacionada con la variable dependiente o independiente pero que no forma parte del experimento.	Variable amenazadora
Moderadora	Variable relacionada con las variables independiente y dependiente y que tiene un impacto en la variable dependiente.	Variable de interacción

Tabla 2.1 Diferentes tipos de variables, sus definiciones y ejemplos.

Variables dependientes

Una **variable dependiente** es la que refleja los resultados de un estudio de investigación. Por ejemplo, si medimos la diferencia entre dos grupos de adultos en cuanto a qué tan bien pueden recordar una serie de 10 dígitos individuales después de un periodo de 5 horas, el número de dígitos recordados es la variable dependiente. Un ejemplo final: si usted está examinando el efecto del grado de participación escolar de los padres sobre las calificaciones de los niños, las calificaciones recibidas por los niños se considerarían una variable dependiente.

Podemos pensar en las variables dependientes como los resultados que podrían depender del tratamiento experimental o de lo que el investigador modifica o manipula.

Variables independientes

Una **variable independiente** representa los tratamientos o condiciones que el investigador controla para probar sus efectos sobre algún resultado. Las variables independientes también se denominan *variables de tratamiento*, y es tal vez dentro de este contexto que más se usa el término. Una variable independiente se manipula en el curso de un experimento a fin de entender los efectos de tal manipulación sobre la variable dependiente.

Por ejemplo, podríamos querer probar la efectividad de tres diferentes programas de lectura sobre las habilidades de lectura de los niños. El diseño se ilustra en la figura 2.1. El método A incluye tutoría. El método B incluye tutoría y recompensas, y el método C no incluye ni tutoría ni recompensas (estos niños simplemente pasan tiempo con el maestro). En este ejemplo se manipula el método para enseñar a leer, y es la variable independiente. Las calificaciones de lectura son la variable resultante, o dependiente. Este experimento incluye tres niveles de una variable independiente y una variable dependiente.

Método para enseñar a leer (variable independiente)

Método A (con tutorial)	Método B (con tutoría y recompensas)	Método C (sin tutoría ni recompensas)
Calificaciones de lectura (variable dependiente)	Calificaciones de lectura (variable dependiente)	Calificaciones de lectura (variable dependiente)

Figura 2.1 Los diseños de investigación pueden adoptar muchas formas diferentes. Aquí, el investigador está examinando los efectos de tres diferentes métodos para enseñar a leer sobre las calificaciones de lectura. Observe que en el método C no se implementa ninguno de los tratamientos. Esta es la condición de control.

¿Qué tal si quisiéramos investigar si existe o no una diferencia entre hombres y mujeres en sus calificaciones de matemáticas según alguna prueba estandarizada? En este ejemplo, la variable independiente es el género (masculino

o femenino) y la variable resultante o dependiente es la calificación de matemáticas. O bien, podríamos examinar los efectos del número de horas que se ve televisión a la semana (menos de 25 para el grupo A; 250 más para el grupo B) en las habilidades de lenguaje. Aquí, la cantidad de tiempo que se dedica a ver televisión es la variable independiente, y las habilidades de lenguaje son la variable dependiente.

La regla general a seguir es que cuando el investigador manipula algo o asigna participantes a grupos con base en alguna característica, como edad, grupo étnico o tratamiento, esa variable es la variable independiente. Cuando el investigador examina algún resultado para determinar si el agrupamiento tuvo algún efecto, está examinando la variable dependiente. En algunos casos, cuando al investigador no le interesa determinar los efectos de una cosa en otra, sino sólo las relaciones entre variables, no hay variables independientes. Por ejemplo, si sólo nos interesa la relación entre la cantidad de tiempo que un padre dedica a estar con sus hijos y su desempeño en el trabajo, no estamos manipulando nada.

Puesto que las variables independientes deben asumir más de un valor (pues son variables), cada una debe tener por lo menos dos niveles, o valores. Por ejemplo, si un investigador estudia los efectos de las diferencias de género (la variable independiente) sobre el desarrollo del lenguaje (la variable dependiente), la variable independiente tiene dos niveles, masculino y femenino. De forma similar, si un investigador estudia el efecto de las diferencias de edad sobre la tensión para personas con edades de 30-39, 40-49 y 50-59 años, la variable independiente sería la edad, y tendría tres niveles.

¿Qué sucede si tenemos más de una variable independiente? Examine la figura 2.2, que representa un **diseño factorial** en el que el género, la edad y la clase social son variables independientes. Los diseños factoriales son experimentos que incluyen más de una variable independiente. Aquí hay dos niveles de género, tres niveles de edad y tres niveles de clase social, para dar un diseño de 2 por 3 por 3, o un total de 18 (!) combinaciones distintas, o **celdas**, de niveles de variables independientes. Es evidente que a medida que se añaden variables independientes a un diseño de investigación, el número total de celdas puede aumentar rápidamente.

Edad	Género			Femenino			
	Clase social	Baja	Media	Masculino	Baja	Media	Alto
1							
30-39							
40-49							
50-59							

Figura 2.2 Muchos experimentos emplean más de una variable independiente. En este ejemplo hay tres: género, clase social y edad.

La relación entre variables independientes y dependientes

¿Con respecto a qué son independientes las variables del mismo nombre? La mejor variable de este tipo es independiente de cualquier otra variable del mismo estudio. Así, la variable independiente puede contribuir con el máximo de información, más allá de lo que pueden ofrecer otras variables independientes.

¿De qué dependen las variables dependientes? La mejor variable dependiente es sensible a cambios en los diferentes niveles de la variable independiente. Si no es así, aunque el tratamiento haya tenido algún efecto, nunca se podrá detectar.

La ecuación siguiente resume la relación entre las variables independientes y dependientes.

$$DV = f(IV1, IV2, \dots, IVk)$$

donde DV = la variable dependiente

f = función de

IV = la(s) variable(s) independiente(s)

Por ejemplo, si queremos investigar los efectos de diferentes técnicas de selección y los años de experiencia sobre la satisfacción en el trabajo, las variables independientes son la técnica de selección (IV_1) y los años de experiencia (IV_2). La variable dependiente (DV) es la satisfacción con el trabajo.

Otros tipos de variables importantes

Las variables independientes y dependientes son los dos tipos de variables que manejaremos con mayor frecuencia en este libro. Sin embargo, existen otras variables que es importante que conozca, porque el entendimiento de qué son y cómo embonan en el proceso de investigación es indispensable para que usted sea un consumidor inteligente y tenga bases firmes como productor incipiente de investigaciones. He aquí otros tres tipos de variables con los que deberá familiarizarse. Recuerde que la tabla 2.1 presenta un resumen de todas las variables que cubriremos en esta sección, e incluye un resumen y un ejemplo de cada una.

Una **variable de control** es una que podría tener influencia sobre la variable dependiente. Por tanto, es necesario eliminar o controlar tal influencia. Por ejemplo, si nos interesa examinar la relación entre la velocidad de lectura y la comprensión de la lectura. Podríamos querer controlar las diferencias en inteligencia, ya que ésta tiene relación tanto con la velocidad de lectura como con la comprensión de la lectura. Así, es preciso mantener constante la inteligencia para tener una buena idea de la naturaleza de la relación entre las variables de interés.

Una **variable extraña** es aquella que tiene un impacto impredecible sobre la variable dependiente. Por ejemplo, si nos interesa examinar los efectos de ver televisión en el aprovechamiento, podríamos encontrar que el tipo de programas de televisión que se miran es una variable extraña que podría afectar el aprovechamiento, ya que programas y documentales de tipo educativo o científico podrían tener un impacto positivo sobre el aprovechamiento, mientras que otros programas podrían tener un efecto negativo.

Una **variable moderadora** es aquella que está relacionada con las variables de interés (como la variable dependiente y la independiente) y enmascara la verdadera relación entre la variable independiente y la dependiente. Por ejemplo, si estamos examinando la relación entre la tasa de delitos y el consumo de helado, más vale que incluyamos la temperatura, que modera esa relación. De lo contrario, nuestras conclusiones serían inexactas.

Hipótesis

En el primer capítulo definimos una hipótesis como "una conjetura educada". Si bien una hipótesis refleja muchas otras cosas, tal vez su papel más importante es reflejar el planteamiento general del problema o pregunta que motivó que se emprendiera el estudio de investigación. Es por ello que es de vital importancia plantear esa pregunta inicial con cuidado y detenimiento, pues nos guiará durante la creación de una hipótesis, la que a su vez ayudará a determinar los tipos de técnicas que usaremos para probar la hipótesis y contestar la pregunta original.

Por ejemplo, la etapa de "¿Qué tal si...?" se convierte en la etapa de planteamiento del problema, que luego conduce a la hipótesis del estudio. He aquí un ejemplo de cada etapa.

Etapa	Ejemplo
"¿Qué tal si...?" :	Me parece que hay varias cosas que podrían hacerse para ayudar a nuestros empleados a disminuir su elevada tasa de ausencias. Después de hablar con varios de ellos me di cuenta de que les preocupa el cuidado de sus niños después de la escuela. ¿Qué tal si se iniciara un programa aquí en la fábrica que pudiera proporcionar supervisión y actividades para los niños?

La hipótesis : Los padres que inscriben sus hijos en programas después de la escuela faltan menos días al trabajo en un año y tienen una actitud más positiva hacia el trabajo, medida según alguna técnica de encuesta laboral estandarizada, que los padres que no inscriben a sus hijos en tales programas.

Así, una buena hipótesis traduce el planteamiento de un problema a una forma más susceptible de probarse empleando los métodos de investigación que tratamos en este libro. Hablaremos de qué hace que una hipótesis sea buena después de definir los dos tipos de hipótesis, la hipótesis nula y la hipótesis de investigación, y cómo se usan.

La hipótesis nula

La hipótesis nula es un bicho interesante. Si pudiera hablar, nos diría algo así como: "Represento la ausencia de una relación entre las variables que estás estudiando". En otras palabras, las hipótesis nulas son expresiones de *igualdad*, como ilustran las hipótesis nulas siguientes:

- No habrá diferencias en la calificación promedio de los alumnos de tercero de secundaria y la de los alumnos de

tercer año de bachillerato en la prueba de memoria ABC.

- No existe relación entre el tipo de personalidad y el éxito en el trabajo.
- No hay diferencias en los patrones de voto en función del partido político.
- La marca de helado preferida es independiente de la edad, género e ingresos del comprador.

Lo que estas cuatro hipótesis nulas tienen en común es que dicen que dos o más cosas son iguales y no están relacionadas entre si.

¿Para qué sirve básicamente una hipótesis nula? Esta hipótesis actúa como *punto de partida* y también como *marca de referencia* contra la cual se medirán los resultados reales de un estudio. Examinemos cada uno de estos fines.

En primer lugar, la hipótesis nula actúa como punto de partida porque es la situación que se acepta como cierta en ausencia de otra información. Por ejemplo, examinemos la primera de las hipótesis nulas anteriores: *No habrá diferencias en las calificaciones promedio de los alumnos de tercero de secundaria y las de los alumnos de tercer año de bachillerato en la prueba de memoria ABC*. Si no tenemos un conocimiento ulterior de las habilidades de memoria de los alumnos de tercero de secundaria y de tercero de bachillerato, no tendremos una razón para creer que habrá diferencias entre los dos grupos. Podríamos especular acerca de por qué uno de los grupos podría tener un mejor desempeño, pero si no tenemos pruebas *a priori* (antes del hecho) no tenemos más opción que suponer que ambos grupos son iguales. Esta falta de relación, a menos que se demuestre otra cosa, es una característica distintiva del método que estamos estudiando. En otras palabras, hasta que no se demuestre que existe una diferencia, hay que suponer que no hay diferencia.

Por añadidura, si hay diferencias entre los dos grupos, tenemos que suponer que las diferencias se deben a la explicación más atractiva para las diferencias entre grupos respecto a cualquier variable: *¡el azar!* Así es; si no contamos con otra información, el azar siempre es la explicación más probable para las diferencias entre dos grupos. Y ¿qué es el azar? Es la variabilidad aleatoria que se introduce en todo estudio en función de los individuos que participan en él y de muchos factores imprevistos, como la forma en que se mide el comportamiento. Por ejemplo, podríamos tomar un grupo de jugadores de fútbol y de básquetbol y comparar las velocidades con que corren. Pero, ¿cómo vamos a saber si algunos jugadores de fútbol practican más, o si algunos jugadores de básquetbol son más fuertes, o si ambos grupos están recibiendo entrenamiento adicional? Es más, tal vez la forma en que se está midiendo su velocidad deja lugar para el azar; un cronómetro defectuoso o un día con viento pueden contribuir a diferencias que no tienen relación con la verdadera velocidad de carrera. Como buenos investigadores, nuestra tarea es eliminar el azar como factor y evaluar otros factores que podrían contribuir a las diferencias entre grupos, como los que se identifican como variables independientes.

El segundo propósito de la hipótesis nula es tener una *marca de referencia* con la cual comparar los resultados observados para ver si las diferencias se deben al azar o a algún otro factor. La hipótesis nula ayuda a definir un intervalo dentro del cual cualesquier diferencias que se observen entre los grupos se pueden atribuir al azar (que es lo que propone la hipótesis nula) o a otra cosa (que tal vez sería el resultado de la manipulación de la variable independiente).

La mayor parte de los estudios correlacionales, cuasiexperimentales y experimentales tienen una hipótesis nula implícita; no así los estudios descriptivos e históricos. Por ejemplo, si nos interesa el incremento en el índice de vacunación durante los últimos 70 años (histórico) o qué opina la gente acerca de las becas escolares (descriptivo), probablemente no nos interesan las diferencias entre grupos.

La hipótesis de investigación

Mientras que una hipótesis nula es una expresión de la ausencia de relación entre variables, una **hipótesis de investigación** es una expresión definida de la relación entre dos variables. Por ejemplo, para cada una de las hipótesis nulas que planteamos antes, presentaremos ahora una hipótesis de investigación correspondiente. Tome nota de que dije una y no "la" hipótesis de investigación correspondiente, pues ciertamente puede haber más de una hipótesis de investigación para cualquier hipótesis nula. He aquí algunas hipótesis de investigación que corresponden a las hipótesis nulas antes mencionadas.

- La calificación promedio de los alumnos de tercero de secundaria es diferente de la de los alumnos de tercer año de bachillerato en la prueba de memoria ABC.
- Hay una relación entre el tipo de personalidad y el éxito en el trabajo.

- Los patrones de voto son función del partido político.
- La marca de helado preferida está relacionada con la edad, el género y los ingresos del comprador.

Todas estas hipótesis de investigación tienen algo en común: son expresiones de *desigualdad*. Las hipótesis plantean una relación entre variables y no una igualdad, como hace la hipótesis nula. La naturaleza de esta desigualdad puede adoptar dos formas distintas: *direccional* y *no direccional*. Si la hipótesis de investigación no plantea un sentido para la desigualdad (sólo dice que hay diferencia), es no direccional; si plantea un sentido para la desigualdad (como más que o menos que), se trata de una hipótesis de investigación direccional.

La hipótesis de investigación no direccional

Las **hipótesis de investigación no direccionales** reflejan una diferencia entre grupos, pero no se especifica la dirección de la diferencia. Por ejemplo, la hipótesis de investigación *La calificación promedio de los alumnos de tercero de secundaria es diferente de la de los alumnos de tercer año de bachillerato en la prueba de memoria ABC* es no direccional porque no se especifica el sentido de la diferencia entre los dos grupos. La hipótesis sólo dice que hay una diferencia, pero no dice nada acerca de la dirección de tal diferencia; es una hipótesis de investigación porque se postula una diferencia, pero no se especifica su naturaleza.

Una hipótesis de investigación no direccional como la que describimos aquí se representaría con la ecuación siguiente:

$$H_1: X_9 \neq X_{12}$$

donde H_1 : representa la primera (de posiblemente varias) hipótesis de investigación

X_9 = representa la calificación promedio de la muestra de alumnos de tercero de secundaria en la prueba de memoria

X_{12} = representa la calificación promedio de la muestra de alumnos de tercer año de bachillerato en la prueba de memoria

\neq = indica desigualdad

La hipótesis de investigación direccional

Las hipótesis **de investigación direccionales** reflejan una diferencia entre grupos, y se especifica la dirección de la diferencia.

Por ejemplo, la hipótesis de investigación *La calificación promedio de los alumnos de tercer año de bachillerato es mayor que la de los alumnos de tercero de secundaria en la prueba de memoria ABC* es direccional, pues se especifica el sentido de la diferencia entre los dos grupos. Se postula que una calificación promedio es mayor que la otra.

Las hipótesis direccionales pueden adoptar las formas

-A es mayor que B ($\circ A > B$), o

-A es menor que B ($\circ A < B$)

Ambas representan desigualdades. Una hipótesis de investigación direccional como la que acabamos de describir, en la que se postula que los alumnos de tercer año de bachillerato obtienen en promedio una mejor calificación que los de tercero de secundaria, se representaría con la siguiente ecuación:

$$H_1: X_9 < X_{12}$$

donde H_1 : representa la primera (de posiblemente varias) hipótesis de investigación

X_9 = representa la calificación promedio de la muestra de alumnos de tercero de secundaria en la prueba de memoria

X_{12} = representa la calificación promedio de la muestra de alumnos de tercer año de bachillerato en la prueba de memoria

$<$ = indica menor que

¿Para qué sirve la hipótesis de investigación? Es esta hipótesis la que se prueba directamente en un paso del proceso de investigación. Los resultados de esta prueba se comparan con lo que se esperaría exclusivamente del azar (que refleja la hipótesis nula) para determinar cuál de las dos es la explicación más atractiva de cualesquier diferencias entre grupos que pudiéramos observar.

Diferencias entre la hipótesis nula y la hipótesis de investigación

Además de que la hipótesis nula representa una igualdad y la hipótesis de investigación representa una desigualdad, hay otras diferencias importantes entre los dos tipos de hipótesis. En primer lugar, una (la hipótesis nula) dice que no existe relación entre ciertas variables (igualdad) mientras que la otra (la hipótesis de investigación) dice que existe una relación (una desigualdad). Ésta es la diferencia primaria.

En segundo lugar, las hipótesis nulas siempre se refieren a la *población*, mientras que las hipótesis de investigación siempre se refieren a la *muestra*. Como veremos más adelante en este mismo capítulo, los investigadores escogen una muestra de participantes de una población mucho más grande. Una razón para usar este método es que resulta demasiado costoso y a menudo imposible trabajar con toda la población.

En tercer lugar, dado que no se puede someter toda la población a la prueba (por ser poco práctico, costoso y a menudo imposible), nunca podemos asegurar que no existe una diferencia *real* entre grupos en cuanto a una variable dependiente dada (si aceptamos la hipótesis nula). Más bien, tenemos que inferirlo (indirectamente) de los resultados de la prueba de la hipótesis de investigación, que se basa en la muestra. Por tanto, la hipótesis nula se debe probar en forma indirecta, mientras que la de investigación se prueba directamente.

En cuarto lugar, las hipótesis nulas siempre se expresan usando símbolos griegos, y las de investigación siempre se representan con símbolos romanos. Por ejemplo, la hipótesis nula de que la calificación promedio de los alumnos de tercero de secundaria es igual a la de los de tercero de bachillerato; se representa así:

$$H_0: \mu_9 = \mu_{12}$$

donde H_0 : representa la hipótesis nula

μ_9 = representa el promedio teórico para la población de alumnos de tercero de secundaria

μ_{12} = representa el promedio teórico para la población de alumnos de tercero de bachillerato

Las hipótesis de investigación se expresan empleando símbolos romanos como X para representar el promedio de una muestra.

Por ejemplo, la hipótesis de investigación de que la calificación promedio de una muestra de alumnos de tercero de secundaria es menor que la de una muestra de alumnos de tercero de bachillerato se representaría con la ecuación siguiente:

$$H_1: X_9 < X_{12}$$

donde H_1 : representa la hipótesis de investigación

X_9 = representa el promedio para una muestra de alumnos de tercero de secundaria

X_{12} = representa el promedio para una muestra de alumnos de tercero de bachillerato

Por último, dado que no podemos probar directamente la hipótesis nula, se trata de una hipótesis *implícita*. La hipótesis de investigación es *explícita*. Es por esta razón que casi nunca se plantean hipótesis nulas en los informes de investigación, aunque casi siempre vemos un planteamiento de la hipótesis de investigación.

¿Qué hace que una hipótesis sea buena?

Ahora sabemos que las hipótesis son conjeturas educadas. A igual que todas las conjeturas, algunas son mejores que otras desde el principio. No puedo hacer demasiado hincapié en lo importante que es hacer la pregunta cuya respuesta se desea y tener presente que cualquier hipótesis que presentemos es una extensión directa de la pregunta original que hicimos. Esta pregunta refleja nuestros intereses personales y las investigaciones que se han efectuado anteriormente.

Con esto en mente, he aquí algunos criterios que podríamos usar para decidir si una hipótesis que leemos en un

informe de investigación o las que nosotros mismos formulamos son aceptables.

Como ilustración, utilicemos un ejemplo de un estudio que examina los efectos de los programas de guardería después de la escuela para empleados que trabajan tarde, sobre la forma como los padres se ajustan al trabajo. He aquí una hipótesis bien escrita:

Los padres que inscriben a sus hijos en programas después de la escuela faltaran menos al trabajo en un año y tendrán una actitud más positiva hacia su actividad laboral, medida según la encuesta de Actitud Hacia el Trabajo (AHT), que los padres que no inscriben a sus hijos en tales programas.

He aquí los criterios.

Primero, *una buena hipótesis se expresa en forma declarativa y no como pregunta*. En el ejemplo anterior no se planteó: "¿Cree usted que los padres y las compañías para las que trabajan se beneficiarán...?", pues las hipótesis son más efectivas cuando se expresan como una afirmación clara y categórica.

Segundo, *una buena hipótesis plantea una relación esperada entre variables*. La hipótesis que estamos usando como ejemplo describe claramente la relación entre el cuidado de los niños después de la escuela, la actitud de los padres y la tasa de ausencias. Estas variables se están probando para ver si una de ellas (inscripción en el programa después de la escuela) afecta las otras (tasa de ausencias y actitud).

¿Se percató de la palabra *esperada* en el criterio anterior? Se define una relación esperada con el fin de evitar el "viaje de pesca" (también conocido como estrategia de "escopeta") que podría ser tentador emprender pero que no resulta muy productivo. La estrategia de viaje de pesca es aquella en la que se lanza el sedal con un anzuelo y se toma cualquier cosa que muerde. Se recopilan datos sobre todo lo que se puede sin importar qué intereses tengamos y sin pensar siquiera en si la recopilación de tales datos es una parte razonable de una investigación científica. Es como cargar las escopetas y dispararle a todo lo que se mueva. Es probable que le atinemos a algo. El problema es que tal vez aquello a lo que le atinamos no sea lo que queremos y, lo que es peor, que fallemos en atinarle a lo que sí queremos o, peor aún (si es posible), ¡que ni siquiera sepamos a qué le atinamos!

Los buenos investigadores no buscan cualquier cosa que puedan atrapar o derribar; quieren resultados específicos. Para obtenerlos, los investigadores necesitan que sus preguntas iniciales y sus hipótesis sean claras, categóricas y fáciles de entender.

Tercero, *las hipótesis reflejan la teoría o los estudios publicados (bibliografía) en los que se basan*. Como vimos en el capítulo 1, los logros de los científicos casi nunca pueden atribuirse únicamente a su intenso trabajo; también se deben a que muchos otros investigadores vinieron antes y establecieron los cimientos para exploraciones posteriores. Una buena hipótesis refleja esto, en cuanto a que tiene un vínculo sustancial con los estudios publicados (la bibliografía) y las teorías existentes. En el ejemplo anterior, supongamos que hay trabajos publicados que indican que los padres se sienten más a gusto sabiendo que sus hijos están recibiendo atención en un entorno estructurado, y así pueden ser más productivos en su trabajo. Saber esto nos permitiría plantear la hipótesis de que un programa después de la escuela proporcionaría a los padres la seguridad que están buscando, y a su vez les permitiría concentrarse en el trabajo y no en el teléfono para averiguar si Raquel o Gregorio llegó sano y salvo a casa.

Cuarto, *una hipótesis debe ser breve y concisa*. Queremos que la hipótesis describa la relación entre las variables en una forma declarativa y que sea lo más concisa posible. Cuanto más concisa, más fácil será para otros (¡como los miembros del comité para su tesis de maestría!) leer su investigación y entender exactamente qué está proponiendo y cuáles son las variables importantes. De hecho, cuando la gente lee y evalúa investigaciones (de lo cual hablaremos más adelante en este capítulo), lo primero que hace en muchos casos es encontrar las hipótesis para tener una buena idea de cuál es el propósito general de la investigación harán las cosas. Una buena hipótesis nos dice estas dos cosas.

Quinto, *las buenas hipótesis se pueden probar*. Esto significa que en realidad se puede poner en práctica la intención de la pregunta que se refleja en la hipótesis. En la hipótesis por ejemplo podemos ver que la comparación importante es entre los padres que han inscrito a su hijo en un programa para después de la escuela y aquellos que no lo han hecho. La actitud hacia el trabajo y las ausencias. Las dos cosas son objetivos razonables, la actitud se mide según la encuesta de Actitud Hacia el Trabajo (un título ficticio para tener una idea), y el ausentismo (el número de días que el empleado faltó al trabajo) es una medida fácil de registrar y que no admite ambigüedades. Pensemos cuánto más difíciles serían las cosas si la hipótesis se expresara de esta manera:

*Los padres que inscriben a su hijos en guarderías después de la escuela se sienten mejor respecto al trabajo” si bien es posible que esto comunique el mismo mensaje, los resultados podrían ser más difíciles de interpretar dada la naturaleza ambigua de palabras como *se sienten mejor*.*

En síntesis las hipótesis deben

- expresarse en forma declarativa,
- postular una relación entre variables,
- reflejar una teoría o cuerpo de bibliografía sobre el que se basan
- ser breves y concisas, y
- poder probarse

Si una hipótesis satisface estos cinco criterios, sabemos que es lo bastante buena como para continuar un estudio que pruebe con exactitud la pregunta general de la cual se derivó la hipótesis

Muestras y Poblaciones

Como buen científico a usted le gustaría poder decir que si *el método A es mejor que el método B es y será cierto eternamente y siempre para toda la gente*. De hecho, si usted investigara lo suficiente acerca de los méritos relativos de los *métodos* A y B y prueba suficientes personas es posible que algún día pueda decir eso, aunque no es verosímil. Se requiere demasiado dinero y tiempo (¡toda esa gente!) para realizar tanta investigación.

Dadas las restricciones, bajo las cuales viven casi todos los científicos, de que nunca hay suficientes fondos ni tiempo para la investigación, la siguiente mejor estrategia es tomar una porción de un grupo mayor de participantes una parte de 3 un grupo mayor de participantes y realizar la investigación con ese grupo menor, en ese contexto, el grupo mayor se llama **población** y el más pequeño, se llama **muestra**.

Deben seleccionarse muestras de poblaciones a modo de maximizar la probabilidad de que la muestra represente la población de la forma más fiel posible. El objetivo es tener una muestra que se parezca lo más posible a la población. La implicación más importante es asegurar la similitud entre las dos es que una vez que se finalice la investigación los resultados basados en la muestra se puedan generalizar a la población. Si la muestra representa fielmente a la población, se dice que los resultados del estudio son **generalizables o que tienen generalizabilidad**. Hablaremos mas acerca de los distintos tipos de procedimientos de muestreo en el capítulo 4.

El Concepto de significancia

Es probable que no haya un término o concepto que produzca más confusión en el estudiante que el de significancia (significatividad) estadística. Explicaremos con detalle este concepto en el capítulo 8, pero es importante presentar al lector este término lo antes posible en el libro porque es un componente básico e importante del entendimiento del proceso de investigación.

Al principio de este capítulo usted leyó una sencilla reseña de un estudio en el que dos investigadores examinaron las diferencias entre adolescentes cuyas madres trabajan y adolescentes cuyas madres no trabajan (además de la situación familiar, pero para este ejemplo nos limitaremos a los grupos de trabaja y no trabaja). Modifiquemos el significado de la palabra *diferencias* por medio del adjetivo *significativas*. Lo que queremos decir con *diferencias significativas* es que las diferencias que se observan entre los adolescentes cuyas madres trabajan y aquellos cuyas madres no trabajan se deben a alguna influencia y no aparecen simplemente por casualidad. En el presente ejemplo, dicha influencia o factor es el hecho de que la madre trabaje o no trabaje. Supongamos que hemos controlado todos los demás factores que podrían causar alguna diferencia. Entonces, lo único que puede explicar las diferencias entre los dos grupos de adolescentes es el hecho de que sus madres trabajen o no. ¿Correcto? Sí. ¿Ya terminamos? todavía no.

Dado que el mundo y usted y yo y el proceso de investigación no somos perfectos, es preciso considerar cierto margen. En otras palabras, necesitamos poder decir que, si bien estamos bastante seguros de que la diferencia entre los dos grupos de adolescentes se debe a que la madre trabaje o no, no podemos estar absolutamente, 100%, positivamente, inequívocamente, indiscutiblemente (¿capta la idea?) seguros.

¿Por qué? Por muchas razones distintas. Por ejemplo, podría ser (Dios no lo quiera) que simplemente estemos equivocados. Quizá, durante este experimento en particular, las diferencias no se debieron al grupo al que pertenecen

los adolescentes sino a algún otro factor que por descuido no se tomó en cuenta, como las experiencias fuera del hogar. ¿Qué tal silos miembros de un grupo eran en su mayoría hombres y reaccionaron de manera muy diferente de como lo hicieron los miembros del otro grupo, que eran en su mayoría mujeres? Si usted es un buen investigador y se prepara a conciencia, tales diferencias entre grupos son resultados poco probables pero no por ello imposibles. Este factor (género) y otros ciertamente podrían tener un impacto sobre la variable resultante o independiente, y a su vez sobre los resultados finales y las conclusiones a las que usted haya llegado.

¿Qué hacer, entonces? En la mayor parte de las empresas científicas que implican proponer hipótesis y examinar diferencias entre grupos, con toda seguridad habrá una cierta cantidad de error que es imposible controlar. El **nivel de significancia** es el riesgo asociado a no tener una certeza de 100% de que la diferencia se debe a lo que creemos que se debe, pues podría deberse a algún factor imprevisto. Si vemos que un estudio produjo hallazgos *significativos* en el nivel de 0.05, la traducción es que existe una posibilidad en 20 (que equivale a 0.05, o sea 5%) de que las diferencias observadas no se hayan debido a la razón postulada en la hipótesis (la variable independiente) sino a alguna o algunas otras razones desconocidas. Su tarea como buen científico es reducir esta posibilidad hasta donde sea posible eliminando todas las razones que podrían competir, distintas de la que se está probando, para cualesquier diferencias que usted haya observado. Puesto que no es posible eliminar *plenamente* dicha posibilidad, se le maneja asignándole un nivel de probabilidad e informando los resultados con esa salvedad. La especificación de niveles de significancia determinados tiene un aspecto técnico, que veremos más a fondo en el capítulo 8.

Muy bien, con esto ya tenemos un vocabulario y conocimientos básicos para entender la mayor parte de los términos que se emplean en el proceso de investigación. A familiarizarse con ellos usted habrá establecido los cimientos que le permitirán continuar con la siguiente sección de este capítulo, que se ocupa de la forma de leer y evaluar artículos de investigación.

Como leer y evaluar investigaciones

Casi todas las actividades de investigación en las que usted participará implicarán leer artículos de investigación que aparecen en publicaciones periódicas y libros. De hecho, uno de los defectos más comunes de quienes se inician en el campo de la investigación es no estar suficientemente familiarizado con la abundancia de informes de investigación en su área de interés específica. Sería muy inusitado encontrar un tema de investigación sobre el cual o en relación con el cual no se haya hecho nada. Es posible que usted no logre encontrar algo que trate exactamente el tema que desea indagar (como los cambios en el comportamiento de los adolescentes australianos que viven en el desierto), pero hay abundante información acerca del comportamiento adolescente y mucho sobre los niños que viven en Australia. Parte de su trabajo como buen científico es presentar argumentos que destaquen la importancia de estudiar estos factores. Esto puede hacerse leyendo y evaluando investigaciones realizadas en diferentes disciplinas sobre el mismo tema.

¿Qué aspecto tiene un artículo de investigación?

La única forma de adquirir experiencia en el entendimiento de los resultados de estudios de investigación es leer los informes correspondientes y tratar de entender qué están diciendo. Comience con una publicación en su propia área. ¿No sabe de ninguna? Entonces haga una de dos cosas:

- Haga una cita con un profesor del área en la que usted está interesado y pregúntele "¿Cuál es la mejor publicación de investigación en mi área?"
- visite la biblioteca y examine el índice de publicaciones periódicas. Lo más seguro es que encuentre decenas, si no cientos, de publicaciones.

Por ejemplo, para quienes están interesados en educación y psicología y áreas afines, he aquí una muestra de 10 publicaciones de investigación que fueron calificadas por 700 personas como aquellas en las que más les gustaría publicar sus trabajos y las que encontrarían más útiles para informar de hallazgos de investigación importantes (Terrance y Johnson, 1978). Si estos 700 investigadores afamados consideran valiosas estas fuentes de información, ¿no serían un lugar excelente para que usted comience?

- *American Educational Research Journal*
- *American Psychologist*
- *Educational Researcher*
- *Educational and Psychological Measurement*
- *Harvard Educational Review*

- *Journal of Educational Research*
- *Journal of Educational Psychology*
- *Journal of Educational Measurement*
- *Phi Delta Kappan*
- *Review of Educational Research*

He aquí otros 10 que se concentran primordialmente en psicología:

- *Child Development*
- *Cognition*
- *Human Development*
- *Journal of Experimental Psychology*
- *Journal of Personality and Social Psychology*
- *Journal of Applied Developmental Psychology*
- *Pediatrics*
- *Perceptual and Motor Skills*
- *Psychological Bulletin*
- *Sex Roles*

El lector encontrará una lista mucho más amplia en el siguiente capítulo.

Criterios para juzgar un estudio de investigación

Nunca es fácil juzgar el trabajo de otra persona. Un buen punto de partida podría ser la lista de verificación siguiente, que está organizada con el fin de ayudarle a concentrarse en las características más importantes de cualquier artículo publicado. Estas ocho áreas ayudan a entender mejor el formato general de este tipo de informes y a estimar qué tan bien el autor o los autores comunican lo que hicieron, por qué lo efectuaron y qué significado tiene lo que realizaron.

1. La revisión de investigaciones previas

Los trabajos publicados que se citan en el estudio, ¿qué tan cerca se vinculan con la bibliografía previa?

¿Es reciente la reseña?

¿Sabe usted de alguna referencia seminal o sobresaliente que se haya omitido?

2. El problema y el propósito

¿Puede usted entender el planteamiento del problema? ¿Se expresa claramente el propósito del estudio?

¿El propósito parece estar vinculado con la bibliografía reseñada? ¿Se expresa claramente el objetivo del estudio?

¿Hay una justificación conceptual en la que se basen las hipótesis? ¿Se explica por qué es importante realizar el estudio?

3. La hipótesis

¿Se expresan claramente las hipótesis de investigación?

¿Se expresan explícitamente las hipótesis de investigación?

¿Las hipótesis expresan una asociación clara entre variables?

¿Las hipótesis se basan en una teoría o en una reseña y presentación de bibliografía pertinente? ¿Se pueden probar las hipótesis?

4. El método

¿Se definen claramente las variables independientes y dependientes? ¿Son completas la definición y la descripción de las variables?

¿Queda claro cómo se realizó el estudio?

5. La muestra

¿Se seleccionó la muestra de un modo tal que, en su opinión, es representativa de la población? ¿Queda claro de dónde provino la muestra y cómo se seleccionó?

Los participantes del estudio, ¿qué tan similares son a los que se han empleado en estudios similares?

6. Resultados y discusión

¿El autor relaciona los resultados con la reseña de la bibliografía? ¿Los resultados están relacionados con la hipótesis?

¿La discusión de los resultados es congruente con los resultados?

¿La discusión deja demostrada la verdad o la falsedad de la hipótesis inicial que el autor presentó?

7. Referencias

¿Está actualizada la lista de referencias?

¿Es consistente el formato de las referencias? ¿Son completas las referencias?

¿La lista de referencias refleja las fuentes de referencia más importantes del área?

8. Comentarios generales acerca del informe

¿El informe tiene una redacción clara y es comprensible? El lenguaje empleado, ¿tiene algún sesgo o predisposición?

¿Qué puntos fuertes y débiles tiene la investigación?

¿Cuáles son las implicaciones primarias de la investigación? ¿Qué haría usted para mejorar la investigación?

Principios básicos de La Investigación ética

Si bien los investigadores deben sentir entusiasmo y anticipación por su trabajo, lo más importante que deben tener presente es que los participantes son seres humanos. Es necesario tratar a estas personas de modo tal que siempre mantengan su dignidad a pesar de la investigación y de sus resultados, y esto no es nada fácil.

Las exigencias de la investigación ética del comportamiento han creado todo un campo de estudio que se llama (¡vaya sorpresa!) ética. En tanto los investigadores continúen utilizando a seres humanos y animales como participantes, la forma como tratamos a esas personas y animales, y los beneficios, aunque sean indirectos, que puedan obtener de su participación, son cuestiones críticas que es preciso mantener en primer término entre todas nuestras consideraciones.

Más adelante en este capítulo enumeraremos las pautas específicas publicadas por grupos profesionales para que las acaten sus miembros, pero primero abordemos las cuestiones generales que surgen en cualquier análisis del comportamiento ético.

Mantenimiento de la intimidad

El mantenimiento de la intimidad se relaciona con varias inquietudes, pero lo hace de manera más directa con el anonimato. Ser anónimo dentro de un contexto de investigación implica que no hay manera de que nadie, con excepción del investigador principal (por lo regular el director) pueda vincular los resultados de un experimento con el individuo asociado a tales resultados.

La forma más común de mantener el anonimato es utilizar una sola hoja maestra que contenga el nombre de cada participante junto con su número de sujeto. En las hojas de calificación, hojas de codificación y demás materiales de prueba sólo se anotará el número de sujeto. La lista que vincula los números con los nombres se mantiene en un lugar seguro y privado, a menudo bajo llave.

Una segunda preocupación relacionada con la privacidad es no invadir los espacios privados de otras personas para observar comportamientos y recabar datos. Por ejemplo, no sería ético grabar subrepticamente la interacción verbal entre un psicoterapeuta y sus clientes. Si bien ésta podría ser una fuente de abundante información, no sería legítima a menos que el cliente y el terapeuta estén de acuerdo con ello.

Coacción

Nunca, por ninguna razón, debe obligarse a la gente a participar. Como usted sabe, los estudiantes universitarios y sobre todo los inscritos en cursos de psicología de nivel introductorio suelen ser la población más utilizada en muchos estudios de investigación de diferente naturaleza. ¿Es siempre ético pedir a tales estudiantes que participen

en un experimento? Probablemente no. A pesar de ello, a muchos estudiantes se les exige participar como requisito del curso. Asimismo, es común que se exija a los empleados de una empresa responder encuestas, llenar cuestionarios y proporcionar otros tipos de información con fines de investigación como parte de sus obligaciones relacionadas con el trabajo.

La clave aquí es no obligar a las personas a participar. Si no quieren participar, se les deberá ofrecer una forma alternativa de cumplir con los requisitos de sus cursos o su trabajo.

Consentimiento informado

Es probable que éste sea el requisito más importante, y el formato de consentimiento informado podría ser la herramienta que por sí sola asegure un comportamiento ético. Sin duda, todo proyecto de investigación que utilice participantes humanos debe incluir un formato de consentimiento informado el cual leerá y firmará cada participante o la persona que autorice la participación (en el caso de un niño, el padre firmaría el formato).

¿Qué aspecto tiene un formato de consentimiento? Presentamos uno en la figura 2.3. No se trata únicamente de una invitación a participar (aunque también puede serlo), sino una descripción de lo que sucederá durante el curso de la investigación. Como se dará cuenta, una carta de este tipo debe contener al menos la siguiente información:

- el propósito de la investigación
- la identidad y la filiación del investigador (o los investigadores)
- lo que está haciendo el investigador (o los investigadores)
- cuánto tiempo durará la participación
- el derecho del sujeto a retirarse del estudio en cualquier momento y por cualquier razón
- los posibles beneficios para el individuo y para la sociedad
- los posibles daños, riesgos o molestias para el individuo
- un compromiso de mantener la confidencialidad estricta de los resultados
- cómo obtener una copia de los resultados
- cómo ponerse en contacto con el investigador (o los investigadores) en caso de tener dudas
- un lugar para la firma del sujeto (o su padre) que indique que está de acuerdo en participar y que entiende el propósito de la investigación

La carta está impresa en papel membretado e ilustra todos los puntos anteriores. No está escrita en jerga científica, y es tan sencilla y clara como es posible. El objetivo aquí es informar, no coaccionar ni persuadir a las personas para que participen.

**Departamento de Psicología e Investigación
Educativa
University of Kansas e Lawrence, KS 66045**

13 de noviembre de 1995

Estimados señor y señora Eisner:

El Departamento de Psicología e Investigación Educativa de la University of Kansas apoya la práctica de obtener el consentimiento informado de, y proteger a, los sujetos humanos que participen en investigaciones. La siguiente información tiene por objeto ayudarle a decidir si permitirá que Nikki y Alexandra participen en el presente estudio. Usted está en libertad de retirar a cualquiera de ellas, o a ambas, del estudio en cualquier momento.

Se pedirá a sus hijas que participen en un juego con otro niño que tiene una discapacidad, en un cuarto provisto de juguetes y libros, y se grabará en videocinta el comportamiento de sus hijas. Una sesión durará aproximadamente 25 minutos. Nos interesa estudiar la interacción entre niños que tienen una discapacidad y niños sin discapacidad. Esta información es importante porque nos ayudará a desarrollar métodos que aumenten la eficacia de los esfuerzos por integrar niños con discapacidades en las aulas educativas normales.

Nos gustaría que sus hijas participaran, pero tal participación es estrictamente voluntaria. Le aseguramos que el nombre de sus hijas no se mencionará en los hallazgos de la investigación. La información sólo se identificará mediante un código numérico.

Si le gustaría tener información adicional acerca del estudio antes o después de que finalice,

puede ponerse en contacto con cualquiera de nosotros por teléfono o por correo. Agradecemos su atención y apreciamos su interés y cooperación.

Atentamente,

Bruce Saxon
Estudiante de posgrado
(913)123-4567

Sam Fine
Profesor
(913)123-4567

Firma del padre o tutor

Figura 2.3 Formato de consentimiento representativo que incluye toda la información importante que necesitan tener los posibles participantes para saber exactamente qué sucederá.

Consentimiento informado en el caso de niños

Existe un problema obvio cuando se trata de obtener el consentimiento informado en el caso de niños. Un ejemplo es una investigación del desarrollo visual de niños pequeños, en la que el niño es demasiado joven para dar cualquier tipo de consentimiento. Queda a juicio de los padres decidir si permitirán o no que su hijo participe.

Sin embargo, la ética de la investigación con niños implica una gran cantidad de cuestiones que van mucho más allá del difícil proceso de asegurar que los niños no se expondrán a ningún peligro, sea físico o psicológico. Por ejemplo, ¿los niños de 12 años tienen edad suficiente para tomar la decisión de retirarse del estudio, opción que debe quedar bien clara en el formato de consentimiento? ¿Pueden ellos entender las implicaciones a largo plazo de la investigación en la que están participando? ¿Y los posibles riesgos? Es aquí donde debe entrar en juego el buen juicio del investigador. Si un niño está convencido de que no desea participar, tal vez usted pierda ese sujeto y esos datos, pero es preciso respetar sus deseos tanto como los de un adulto. Además, una participación forzada podría causar el descontento o enojo del niño y dar lugar a datos poco confiables.

A medida que aumenta la edad de los niños, el problema se vuelve más complejo. Por ejemplo, ¿qué hay con el niño de 12 años que tiene edad suficiente para entender el propósito del experimento? ¿Debe él o ella firmar el formato de consentimiento además del padre, la madre o ambos? Ningún investigador cuerdo se abstendría de obtener primero la autorización de los padres o de alguno de ellos. Además, cuando se utilizan niños de edad escolar en investigaciones, las autoridades de educación podrían exigir que un comité de investigación de la escuela revise la propuesta. Por añadidura, cada vez más investigadores en Estados Unidos, por ejemplo, sienten la necesidad de contar con un seguro de responsabilidad para protegerse en caso de que algún padre enojado promueva una demanda o que ocurra un perjuicio no intencional.

El mejor consejo que puedo dar es hacer cualquier sesión experimental o de tratamiento con niños lo más placentera posible. Una forma de hacerlo es tratar de entusiasmarlos, de hacer interesantes las actividades y recompensarlos al terminar (en tanto la promesa de una recompensa no interfiera lo que se está tratando de estudiar). Pero, sobre todo, recuerde que los niños son física, emocional y socialmente diferentes de los adultos, y que hay que tomar en cuenta esas diferencias cuando se les usa como sujetos.

Confidencialidad

El anonimato significa que los registros no pueden vincularse con nombres. La confidencialidad se mantiene cuando cualquier cosa que se averigua acerca del participante se mantiene en el más estricto secreto. Esto implica que, si es necesario, hay que disfrazar la información (lo cual también se relaciona con el anonimato) y, sobre todo, que todos los datos se deben mantener en una situación controlada.

La mejor manera de mantener la confidencialidad es minimizar el número de personas que ven o manejan los datos. No hay mejor ejemplo de esto que las recientes inquietudes acerca del SIDA y los resultados de las pruebas de selección. La gente se muestra renuente a someterse a ensayos de VIH (el virus asociado al SIDA) porque les preocupa que las compañías de seguros y los empleadores potenciales tengan acceso a los resultados de los ensayos y los usen contra el individuo cuando solicite un trabajo o un seguro de salud o de vida.

Protección contra daños

Ante todo, se debe evitar que los sujetos sufran algún daño físico o psicológico. Si antes de comenzar hay alguna duda respecto a la posibilidad de un riesgo significativo (relativo a los beneficios), no se deberá aprobar el experimento. Observe que el foco aquí son los riesgos y los beneficios. En el caso de un niño desahuciado, es posible que las técnicas más drásticas e incluso no confirmadas que podrían salvar la vida del niño (aunque también podrían apresurar su fallecimiento) impliquen un riesgo elevado, pero los beneficios potenciales podrían ser igualmente importantes.

Compartir los resultados

El conocimiento científico pertenece al dominio público y, aunque ha habido discusiones acaloradas acerca de cuándo revelar qué a quiénes, la mayoría de los investigadores aceptan que es importante presentar los descubrimientos nuevos al público tan pronto como resulte práctico y posible. Una vez que usted finaliza su investigación y redacta el informe final, deberá estar dispuesto a compartir sus hallazgos con otros.

Entre los más importantes de esos "otros" están las personas que participaron en su experimento. En términos prácticos, usted podría ofrecerse a enviarles un resumen del informe final o celebrar una reunión en la que se les informe de los resultados.

Comunicación

Otro componente al compartir los resultados de un experimento se presenta cuando es necesario enterar de algo a un grupo dado de sujetos.

Por ejemplo, usted diseña un experimento en el que se miente a un grupo de sujetos como parte del experimento. O bien, usted podría pedir a niños pequeños que no toquen un juguete especialmente atractivo y luego grabar en video su conducta sin su conocimiento.

Una vez finalizado el experimento, es obligación de usted informar a los sujetos que se les ha engañado hasta cierto punto para los fines del experimento. La mayoría de la gente acepta esto de buena gana (como las víctimas de los programas de televisión de "cámara escondida"), pero algunos se molestan cuando se enteran de que se les ha manipulado. Si los sujetos siguen enojados, es difícil hacer algo más que pedirles perdón y tratar de dejar en claro las cosas. La forma más fácil de informar a los participantes es hablar con ellos inmediatamente después de la sesión o enviar un boletín que explique a los participantes la intención general y los resultados del estudio pero omitiendo datos específicos, como nombres.

Distribución de los beneficios

El último principio tal vez sea el que se viola con mayor frecuencia. Veamos esta situación. En un experimento se usó un tratamiento para mejorar la memoria de personas de edad avanzada que están en las primeras etapas del mal de Alzheimer, una enfermedad devastadora y casi siempre mortal. Digamos que el investigador usa dos grupos, uno que recibe el adiestramiento (el grupo experimental) y otro que no lo recibe (el grupo de control). Para beneplácito del investigador; el grupo de tratamiento aprende con mayor rapidez y recuerda muchas más cosas durante un tiempo mucho mayor. ¡Éxito!

¿Qué debe preocuparnos? Sencillamente, que el grupo que no recibió el tratamiento lo reciba ahora. Esto es el proceder correcto. Cuando un grupo se beneficia por la participación en un estudio, cualesquier otros grupos que hayan participado también deberán beneficiarse. Esto no significa que todas las personas de edad avanzada puedan recibir ayuda; tal cosa podría ser impráctica. Pero todos los participantes en el experimento deberán beneficiarse de la misma manera.

Todas estas cuestiones éticas se aplican a los diferentes tipos de métodos de investigación que describiremos en los capítulos 9, 10 y 11 con diferentes grados de importancia. Por ejemplo, no es preciso preocuparse por informar a los participantes cuando se realiza un estudio de caso, ya que no hay ni tratamiento ni engaño. Tampoco es necesario compartir los beneficios en una situación así.

Cómo asegurar un nivel ético elevado

Hay varios pasos que incluso un investigador principiante puede seguir para asegurar el mantenimiento de los

principios éticos que hemos bosquejado. He aquí algunos de los más importantes.

1.-Realice una simulación en computadora en la que se construyan datos y se sometan a los efectos de diversos tratamientos. Por ejemplo, los psicólogos y estadísticos matemáticos a menudo utilizan estudios Monte Carlo para examinar los efectos de un cambio en una variable (como el tamaño de la muestra) sobre otra (como la exactitud de las mediciones). Es posible construir complejos modelos del comportamiento humano y probar con ellos diferentes supuestos para sacar conclusiones acerca de la conducta humana. Si bien este tipo de trabajos es más bien avanzado, nos puede dar una idea de cómo pueden realizarse ciertos experimentos sin que los "participantes" sean algo más que valores generados por una computadora.

2- Si el tratamiento es perjudicial, no se dé por vencido. Trate de encontrar una población que ya haya sido expuesta a los efectos dañinos de alguna variable. Por ejemplo, los miles de niños y mujeres embarazadas que sufrieron desnutrición durante la segunda) Guerra Mundial proporcionaron una muestra valiosísima para estimar los efectos de la desnutrición sobre el desarrollo del feto y el recién nacido, así como los efectos a largo plazo de la desnutrición sobre los niños pequeños. Si bien esto no es agradable, es prácticamente la única forma en que podría realizarse una investigación semejante. Este tipo de investigación se denomina cuasiexperimental y se tratará con mayor detalle en el capítulo 11.

3.-Siempre obtenga el consentimiento informado. Si el tratamiento incluye riesgo, usted debe estar absolutamente seguro de que el participante y otras personas interesadas (padres, parientes) conocen y entienden los riesgos.

4.-Publique todos sus informes usando datos de grupos, no individuos. Esta medida man-tiene la confidencialidad.

5.-Si sospecha que el tratamiento podría tener efectos secundarios negativos, utilice una muestra pequeña y bien informada hasta que pueda expandir el tamaño de la muestra y la envergadura del proyecto. También, asegúrese de obtener la anuencia de su junta de revisión institucional (hablaremos más acerca de esto un poco más adelante).

6-Haga que sus colegas revisen su propuesta y sobre todo sus procedimientos experimentales antes de comenzar. Hágalas la pregunta: "¿Participaría usted sin temor a resultar perjudicado?". Si responden "no", vuelva a estudiar su proyecto.

7-Casi todas las instituciones públicas (como las universidades públicas) y todos los organismos privados (como algunos hospitales y universidades) tienen lo que se conoce como una **junta de revisión institucional**. Tales juntas consisten en un grupo de personas de varias disciplinas (incluidos representantes de la comunidad) que rinden un juicio acerca de si la participación en el estudio es inocua o no. En la University of Kansas, el grupo se llama Advisory Committee on Human Experimentation (ACHE, Comité Asesor sobre Experimentación Humana). Existe una junta de revisión aparte para experimentos en los que se usan animales.

Los grupos suelen reunirse y luego aprobar o vetar el procedimiento (pero no necesariamente el contenido de la investigación) y toman en consideración los aspectos que hemos estudiado. Estos comités por lo regular se reúnen una vez al mes y, si una propuesta que revisan no es aceptable, invitan al investigador a que vuelva a presentarla ajustándose a sus recomendaciones. En la figura 2.4 se presenta una muestra del formato utilizado por el ACHE en la University of Kansas.

El papel de las organizaciones profesionales

Es obligación insoslayable del investigador cuidar que las normas éticas antes analizadas se tengan siempre presentes al realizar cualquier tipo de investigación.

Hay conjuntos de pautas más formalizados publicados por organizaciones profesionales como la American Psychological Association (APA, Asociación Estadounidense de Psicología), la Society for Research in Child Development (SRCD, Sociedad de Investigación sobre Desarrollo Infantil), la American Sociological Association (Asociación Estadounidense de Sociología), la American Educational Research Association (AERA, Asociación Estadounidense de Investigaciones en Educación) y casi cualquier otro grupo profesional en el área de las ciencias sociales o de la conducta. Para ilustrar lo que sugieren tales pautas, haremos un resumen de las que presenta la American Psychological Association (un grupo de unos 25000 profesionales) y la Society for Research in Children Development (un grupo de unos 6000 profesionales).

**SOLICITUD DE APROBACIÓN DE PROYECTO
COMITÉ ASESOR SOBRE EXPERIMENTACIÓN HUMANA**

Llene esta forma a máquina

1. Nombre del (de los) investigador(es) _____
2. Departamento al que pertenece(n) _____
3. Dirección postal en la universidad _____
4. Teléfono(s): (a) Universidad _____ (b) Particular _____
5. Nombre del miembro académico responsable del proyecto _____
6. Tipo de investigador y naturaleza de la actividad. (Marque las categorías apropiadas.)
_____ Docente o empleado de la Kansas University
 Proyecto a presentarse para financiamiento externo; Organismo
 Proyecto a presentarse para financiamiento interno; Fuente _____
 Proyecto sin financiamiento _____ Otro _____
_____ Estudiante de Kansas University
 Posgrado _____ Licenciatura _____ Especial _____
 Tesis _____ Disertación _____
 Proyecto de clase (número y nombre de la clase) _____ Estudio
independiente (nombre del supervisor académico) _____ Otro (especifique) _____
7. Título de la investigación _____

Todas las solicitudes de estudiantes presentadas al ACHE para su revisión deben ir firmadas por todos los investigadores, incluido el miembro académico que supervisará la actividad de investigación.

8. _____ Individuos distintos del profesorado, personal o estudiantes de Kansas University. Identifique los investigadores y el grupo de investigación _____

9. Certificaciones

Estoy familiarizado con las políticas y procedimientos de la University of Kansas en lo concerniente a sujetos humanos de investigación. Comulgo con esas normas y me ajustaré a las políticas y procedimientos de ACHE.

y

Estoy familiarizado con las pautas publicadas para el tratamiento ético de los sujetos, correspondientes a mi campo de estudio específico (por ejemplo, las publicadas por la American Psychological Association, la American Sociological Association, etcétera).

Fecha _____ Fecha _____

Firma _____ Firma _____
Primer investigador Supervisor académico

Figura 2.4 Ejemplo de formato de revisión institucional

Fecha _____ Fecha _____
Firma _____ Firma _____
Segundo investigador Tercer investigador

Investigador principal
ACHE# _____

Título

10. Conteste las preguntas siguientes relacionadas con la actividad de investigación propuesta: La investigación implica:

	Sí	No
a. ¿drogas u otras sustancias controladas?	_____	_____
b. ¿pago a los sujetos por su participación?	_____	_____
c. ¿acceso a sujetos a través de una institución colaboradora	_____	_____
d. ¿sustancias que los sujetos tomarán intemamente o que se les aplicarán en forma externa?	_____	_____
e. ¿aplicar dispositivos mecánicos o eléctricos (por ejemplo, electrodos) a los sujetos?	_____	_____
f. ¿tomar fluidos (por ejemplo, sangre) o tejidos de los sujetos?	_____	_____
g. ¿tensión (fisiológica o psicológica) experimentada por los sujetos?	_____	_____
h. ¿engañar a los sujetos respecto a cualquier aspecto de los fines o los procedimientos?	_____	_____
i. ¿sujetos que se estima tendrán libertad de consentimiento limitada (niños o adultos de lento aprendizaje, enfermos o ancianos)?	_____	_____
j. ¿cualquier procedimiento o actividad que podría poner en riesgo (psicológico, físico o social) a los sujetos?	_____	_____
k. ¿uso de entrevistas, encuestas, cuestionarios, grabaciones de audio o video?	_____	_____
l. ¿recopilación de datos durante más de un año?	_____	_____
m. se dará una copia de la forma de consentimiento a los sujetos	_____	_____

11. Número aproximado de sujetos que participarán en la investigación

Conteste las preguntas siguientes en esta página. No use hojas de continuación. ACRE n procesará las solicitudes que no respeten las limitaciones de páginas. Vea las instrucciones.

12. Propósito(s) del proyecto

13. Describe los sujetos propuestos (edad, sexo, raza u otras características especiales).

14. Describa cómo se seleccionarán los sujetos.

Figura 2.4 Continuación

Pautas éticas de la APA

El primer conjunto de pautas formuladas por un comité de la APA se presentó en 1953. He aquí un resumen de las más recientes.

1. Cuando se planea un estudio, el investigador debe ser el primer y más importante juez de su aceptabilidad ética.
2. Se debe estimar que los sujetos "no corren riesgo" o "corren un riesgo mínimo".
3. El investigador es responsable de asegurar la aplicación de prácticas éticas, incluido el comportamiento de sus asistentes, estudiantes, empleados, colaboradores y cualquier otro participante en el proceso.
4. Se debe llegar a un acuerdo justo y razonable entre el investigador y los sujetos antes del inicio de la investigación.
5. Si es necesario engañar a los sujetos, el investigador debe estar convencido de que se justifica, y se debe incorporar un mecanismo para asegurar que se revele el engaño a los sujetos a la conclusión del estudio.
6. Los investigadores deben respetar la decisión del sujeto de retirarse del estudio y no deben coaccionarlo para convencerlo de que siga participando.
7. Debe hacerse todo lo posible para proteger a los participantes contra daños físicos y psicológicos.
8. Una vez finalizada la investigación, si el participante lo desea, deben compartirse los resultados y debe darse al participante la oportunidad de aclarar cualesquier discrepancias que pudiera percibir.

9. Si la investigación produjera cualquier tipo de daño, el investigador tiene la responsabilidad de corregir el daño.
10. Toda la información obtenida en un estudio de investigación es confidencial.

Pautas éticas de la SRCD

Dado que éste es un grupo dedicado a aprender más acerca del desarrollo de los niños, sus pautas están redactadas con gran precisión, considerando el bienestar de los niños.

1. Los derechos del investigador supeditan a los derechos del niño, sin importar la edad de éste.
2. Todas las cuestiones éticas que rodean al proyecto de investigación son responsabilidad del investigador en jefe.
3. Si hay cambios a los procedimientos aprobados que pudieran afectar la realización ética de la investigación, es necesario consultar con colegas o expertos.
4. Se deberá informar plenamente al niño sobre el proceso de investigación, y contestar todas las preguntas en una forma que el niño pueda entender.
5. Los niños están en libertad de retirarse de la investigación en cualquier momento.
6. Es necesario obtener el consentimiento informado por escrito de los padres, maestros o quienquiera que sea legalmente responsable del bienestar del niño.
7. También debe obtenerse el consentimiento informado de otros que participen en el experimento (como los padres, etc.) además del niño individual.
8. Se deben dejar muy claras las responsabilidades del niño y del investigador.
9. Si existe alguna posibilidad de causar daños, el investigador debe encontrar una forma alternativa de obtener la información necesaria o debe abandonar la investigación.
10. Si es necesario practicar algún engaño, un comité de colegas del investigador deberá aprobar los métodos propuestos.
11. Toda la información es confidencial.
12. Si se van a usar expedientes institucionales como fuente de información, se debe obtener la autorización de todas las partes afectadas.
13. Los hallazgos de cualquier estudio deben informarse a los participantes de una forma tal que los entiendan.
14. Los investigadores deben tener especial cuidado en cómo informan de los resultados a los niños y nunca deben presentar los resultados en forma de sugerencias.
15. Si durante el curso de la investigación surge información que es importante para el bienestar del niño, el investigador tiene la obligación de proporcionar esa información a los padres, maestros u otras partes según sea apropiado.
16. Deben corregirse todas las consecuencias indeseables.
17. Los investigadores deben ser conscientes de que las investigaciones pueden tener implicaciones políticas, sociales y humanas, y deben tener en cuenta esto al informar y compartir los resultados.
18. Si los tratamientos son efectivos, se deberá ofrecer a los grupos de control una oportunidad similar de recibir el tratamiento.
19. Estas normas éticas se presentarán a los estudiantes en el curso de su entrenamiento.
20. Todos los investigadores tienen la responsabilidad de mantener su propia conducta ética y también la de sus colegas.
21. Los editores de publicaciones que informen de investigaciones hechas con niños deberán proporcionar a los autores espacio para resumir las medidas que tomaron para ajustarse a estas normas. Si no queda claro que éstas se respetaron, los editores deberán solicitar información adicional.
22. Dichas normas siempre están abiertas a discusión y enmienda.

¿Funcionan las normas éticas de la APA y la SRCD? En general, la respuesta es probablemente sí. Sin embargo, dado que toca a cada individuo poner en práctica esas pautas, es difícil tener un panorama objetivo.

Para poder intervenir en el juego de la investigación hay que conocer las reglas. En este capítulo presentamos algunos de los términos y conceptos más importantes relacionados con el proceso de investigación y aquellos que se mencionaran una y otra vez en todo el libro. Cuando comience a revisar la bibliografía, ya sea porque va a preparar una propuesta o por alguna otra actividad de investigación relacionada, vera' cómo se usan estos términos y lo útil que es conocer su significado correcto. También debe tener en mente, antes que nada, los importantes principios de ética básicos que hemos visto mientras planea sus actividades de investigación.

Ejercicios

1. En los ejemplos siguientes, identifique la(s) variable(s) independiente(s) y dependiente(s).
 - a. Dos grupos de niños se sometieron a programas de acondicionamiento físico diferentes para ver si los programas tenían algún efecto sobre su fuerza física.
 - b. Un grupo de 100 fumadores empedernidos se dividió en cinco subgrupos, cada uno de los cuales participó en un programa encaminado a dejar de fumar mediante un método distinto. Después de seis meses de participación en el programa, se contó el número de cigarrillos que cada participante fumó al día.
 - c. A un profesor universitario le interesaba determinar cuál era la mejor manera de enseñar la materia de Introducción a la Psicología y asegurar que sus estudiantes aprendieran el material.

2. ¿Por qué la hipótesis nula siempre es una expresión de igualdad? ¿Por qué la hipótesis de investigación puede asumir varias formas diferentes?

3. Escriba la hipótesis nula y la de investigación para el párrafo siguiente:

Se pidió a un grupo de hombres de mediana edad llenar un cuestionario acerca de sus actitudes hacia el trabajo y la familia. Estos hombres son casados y tienen al menos dos hijos cada uno. Otro grupo de hombres, que no tienen hijos, también contestó el mismo cuestionario.

4. Nadie discute que definir las variables de forma clara e inequívoca es crítico para una buena investigación. Con eso en mente, trabaje en grupo y defina las variables siguientes. Tome nota de cómo las definiciones de las distintas personas reflejan su visión personal de lo que la variable representa, así como de lo fácil que es definir algunas variables y lo difícil que es definir otras.
 - a. inteligencia
 - b. altura
 - c. habilidades sociales
 - d. edad
 - e. agresividad
 - f. conservadurismo
 - g. consumo de alcohol
 - h. espabilamiento
 1. personalidad

Tenga en cuenta que incluso las variables que parecen fáciles de definir; como la altura, pueden asumir diferentes significados y definiciones (alto, 165 cm, imponente).

5. Un investigador dedicó cinco años a un proyecto, y pocos de sus resultados fueron significativos. ¿Cómo puede ser ésta una contribución importante al campo a pesar de la falta de resultados significativos?

6. Indique cuáles de las siguientes son variables y cuáles son constantes:

- a. El color de pelo de Luis
- b. La edad en años
- c. El número de ventanas en el domicilio
- d. El color de un automóvil último modelo
- e. La hora del día
- f. El número de respuestas correctas en el examen de esta semana
- g. El número de firmantes de la Declaración de Independencia
- h. *El nombre de la quinta muchacha de la tercera fila*
- i. La fecha de hoy
- j. El número de palabras recordadas

7. Visite la biblioteca y localice tres artículos científicos en su área de interés. Haga lo siguiente:

- a. Identifique las variables independientes y dependientes.
- b. Especifique cómo se va a medir cada variable dependiente; indique si se define con claridad o en forma confusa.
- c. Para cada variable independiente, identifique el número de niveles de esa variable. ¿Qué otras variables independientes cree usted que sería interesante estudiar?

8. ¿Qué hace que una hipótesis sea buena?
9. ¿Para qué sirve leer y evaluar artículos de investigación?
10. Cite tres de los cinco criterios de una hipótesis aceptable.
11. Cite dos principios básicos de la investigación ética. ¿Por qué son importantes?

¿Quiere saber más?

Si las obras que aquí se recomiendan no están disponibles en las bibliotecas de las instituciones de su comunidad, intente poner en práctica sus habilidades de búsqueda por Internet (véase el capítulo 3).

CAPITULO TRES Selección del problema y revisión de la investigación

En Neil J. Salkind “Métodos de Investigación”, Prentice Hall, México, 1998.

Lo que aprenderá en este capítulo

- Cómo seleccionar un problema de investigación
- Descartar una idea tras otra hasta hallar una que se ajuste al interés propio
- La importancia de la experiencia personal en la selección del problema
- Los pasos de la revisión de la bibliografía
- Las diferentes fuentes de información y cómo usarlas
- Cómo usar revistas, resúmenes bibliográficos e índices
- La diferencia entre recursos primarios y secundarios
- Empleo de una síntesis de la bibliografía
- Cómo funcionan las publicaciones científicas periódicas
- Todo acerca de las búsquedas de bibliografía por computadora

Helo a usted aquí, iniciando un curso que se enfoca en los métodos de investigación, y de repente se le tiene que ocurrir un problema en el que supuestamente usted está interesado. Es probable que usted esté tan ansioso de aprender el material que su profesor expondrá en clase y el contenido de este libro, que apenas tenga tiempo para pensar en otra cosa.

Sin embargo, si se detiene un momento y explora mentalmente algunos de los problemas de las ciencias sociales y del comportamiento que han despertado su interés, sin duda encontrará algo acerca de lo cual usted quiera conocer más. De eso se trata el proceso de investigación: averiguar más acerca de algo que ya se conoce en parte.

Una vez que usted selecciona un área de interés, sólo habrá recorrido parte del camino. Ahora viene la expresión de tal interés en forma de una pregunta de investigación y de una hipótesis formal. El siguiente paso es la revisión de la bibliografía, una frase elegante que al parecer implica que usted va a estar muy ocupado. Una revisión bibliográfica requiere tiempo de biblioteca, toma de apuntes y habilidades de organización, pero proporciona una perspectiva de la pregunta hecha que no puede obtenerse si no se conoce qué otros trabajos se han hecho y qué trabajos nuevos es preciso realizar.

Pero, detengámonos un momento. ¿Cómo se supone que alguien tenga un conocimiento suficientemente amplio del campo como para producir hipótesis bien formuladas antes de revisar la bibliografía y familiarizarse con lo que se ha hecho? Como dijo el poeta John Ciardi, ahí está "el meollo del asunto".

Los filósofos tradicionales e historiadores de la ciencia pretenden hacernos creer que la secuencia de sucesos que conducen a una revisión de lo que se ha hecho antes (y que está revelado en los trabajos publicados) es la que se muestra en la figura 3.1a.

Idea → pregunta de investigación hipótesis de investigación revisión de la bibliografía

Figura 3.1a De la idea a la revisión de la bibliografía (pasando por la hipótesis de investigación).

La secuencia de pasos que aquí se muestra está bien en teoría. Como pronto se dará cuenta usted, el proceso real no es tan lineal como se indica. La pregunta de investigación y la hipótesis de investigación son más bien el resultado de una interacción entre la idea original del científico y una revisión continua y exhaustiva de la bibliografía (¡los buenos científicos leen todo el tiempo!), como se indica en la figura 3.1b. Esto significa que, una vez que se formula una hipótesis, no queda esculpida en piedra, sino que puede alterarse para ajustarla a lo que la revisión de la bibliografía podría reflejar, así como a cualquier cambio que usted pudiera tener en sus ideas.

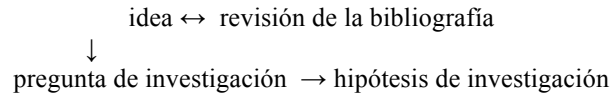


Figura 3.1 b De la idea y la revisión de la bibliografía a la hipótesis.

Por ejemplo, usted podría estar interesado en los efectos de los programas de guardería infantil extendidos después de la escuela, sobre las habilidades de socialización de los niños. Éste es el germen de la idea que usted desea investigar. Una pregunta de investigación podría plantear qué efectos tienen los programas después de la escuela sobre lo bien que los niños se llevan unos con otros. Como hipótesis, usted predice que

Los niños que participan en programas prolongados después de la escuela tienen un nivel de habilidades sociales elevado, medido por la prueba de socialización XYZ.

Usted podría considerar que la hipótesis ya está terminada, pero en realidad su revisión continua de la bibliografía y sus ideas cambiantes acerca de la relación entre las variables influye en la dirección que adoptará su investigación. Por ejemplo, ¿qué tal si usted averigua que ya se realizó un estudio similar que le sugiere añadir una dimensión interesante (como familias con un solo progenitor o con ambos) a su estudio, puesto que la adición es congruente con el propósito del estudio? Usted no tiene por qué restringir su razonamiento creativo ni sus esfuerzos por entender los efectos de estos programas después de la escuela sólo porque ya formuló una hipótesis y llevó a cabo una revisión de las publicaciones. De hecho, una razón para realizar la revisión es ver qué nuevas direcciones podría seguir su trabajo. La revisión de la bibliografía y la idea interactúan para ayudarle a usted a plantear una pregunta de investigación y una hipótesis de investigación pertinentes y conceptualmente sólidas.

En suma, usted casi siempre encontrará que su primer intento de hipótesis necesita revisarse, dado el contenido de la bibliografía revisada. Recuerde, es su idea la que usted explorará. La forma en que la ejecute como estudio de investigación estará determinada por la forma en que plantee la pregunta de investigación y la manera en que pruebe la hipótesis de investigación. Es poco probable que una revisión de la bibliografía pertinente no arroje algo de luz sobre este asunto.

Este capítulo comienza con algunas sugerencias para seleccionar un problema que valga la pena estudiar, y luego se concentra en una descripción de las herramientas y los pasos necesarios para preparar una revisión de la bibliografía.

Selección del problema

La gente asiste a las universidades por diversas razones, que incluyen prepararse para una carrera, lograr las ventajas financieras que una educación puede garantizar, e incluso expandir sus horizontes personales y experimentar el simple gozo de aprender (¡qué idea tan radical!). Muchos de ustedes se inscribieron en este curso específico por una o más de estas razones.

Lo más importante que tienen en común sus clases y sus actividades dentro de este curso es que lo exponen a una gran abundancia de información que de otra manera no experimentarían. Ese es el objetivo primario de dedicar tiempo a seleccionar un problema de investigación que tenga sentido para usted y le interese, y que dé origen a un proyecto que pueda hacer una contribución a su disciplina específica. La selección del problema sobre el que desea trabajar tiene una importancia tremenda por dos razones. Primera, la investigación ocupa una buena cantidad de tiempo y

energía; para hacer esta inversión es crucial el interés personal. Usted trabajará tan arduamente durante todo el proyecto que a veces podría hacerse cuesta arriba continuar los trabajos, incluso si es un proyecto interesantísimo. ¡Imagínese lo que sería si usted no estuviera interesado! Segunda, la selección del problema es sólo el primer paso del proceso de investigación. Si las cosas van bien aquí, los demás pasos, que no son ni más ni menos importantes, tienen una buena oportunidad de dar buen resultado también.

Así como hay diferentes formas de proceder para la selección de un problema de investigación, hay algunos peligros con los que podría toparse. Para que pueda comenzar con el pie derecho, le presentamos una breve reseña de algunos de estos errores que podrían ser fatales.

No es difícil que suceda, pero enamorarse de una idea puede ser fatal. Esto sucede cuando usted se embelesa tanto con su idea y su proyecto, e invierte tanto en él, que no puede soportar hacerle ninguna alteración. Seguramente alguien va a saltar y decir, "¿Qué tiene de malo entusiasmarse con un proyecto?" Mi respuesta es un categórico "Absolutamente nada". Igual que su profesor, la mayoría de los investigadores fomentan y esperan el entusiasmo en los estudiantes (y científicos) como una cualidad importante e indispensable. Pero el entusiasmo no es incompatible con ser objetivo y desapasionado respecto al proceso de investigación en sí (no su contenido). **A veces, y esto sucede sobre todo con los estudiantes que se inician en el campo de la investigación, los investigadores consideran que su pregunta tiene tal magnitud e importancia que se niegan a escuchar a quienes los rodean, incluso a sus asesores, y que están tratando de ayudarles a formular su problema en una forma más precisa y, a la larga, más fácil de abordar. Comprométase con sus ideas y entusiásmese con su tema, pero no tanto que nuble su juicio en cuanto a la forma práctica y correcta de hacer las cosas.**

También, aferrarse a la primera idea que nos viene a la mente no siempre es prudente. Tal vez alguno de ustedes recuerde el personaje de los dibujos animados Betty Boop y su abuelo inventor. Cada vez que Betty tenía un problema, el abuelo se sentaba en su banquillo, cruzaba las piernas (adoptando una pose como la del "Pensador" de Rodin) e ideaba una solución. Como un relámpago del cielo despejado, la bombilla que flotaba sobre su cabeza se encendía y él exclamaba "¡Lo tengo!", pero resultaba que esa todavía no era realmente la solución. Ocurría otro destello, pero ese tampoco era perfecto. Invariablemente, era hasta la tercera vez que se encendía la luz que el abuelo daba en el clavo. ¿Le gusta a usted su primera idea para un estudio de investigación? Magnífico, pero espere un momento antes de correr a poner un anuncio en el periódico solicitando sujetos para la investigación. Tómese unos días para meditar la idea con calma, sin dejar de hablar con otros estudiantes y sus profesores durante esta etapa de maduración. **Las segundas y terceras ideas suelen estar mucho más refinadas y ser más fáciles de poner en práctica y manejables que las primeras. Mientras trabaja, reescriba y refine su trabajo... constantemente.**

¿Quiere la receta para un proyecto condenado al fracaso que no emocione a nadie (salvo, tal vez, usted)? Hacer algo trivial seleccionando un problema sin un fundamento conceptual o sin importancia evidente en el campo puede dar pie a una experiencia frustrante que no le dará satisfacciones. Los estudiantes principiantes que cometen este error a veces adoptan una actitud excesivamente intelectual hacia la importancia de sus planes de investigación y no se detienen un momento para preguntarse "¿Cómo encaja este estudio con lo que se ha hecho antes?" Toda empresa científica tiene como meta más elevada contribuir con información que ayude a entender mejor el mundo en general y el tema de estudio específico en particular. Si usted averigua qué se ha hecho antes leyendo informes de estudios previos y utiliza esa información como base, seguramente ideará un problema de investigación importante y valioso.

Y luego tenemos al **investigador que abarca más de lo que puede apretar**. ¿absurdo? No para los miles de profesores y asesores que se pasan días y días en sus oficinas tratando de convencer a estudiantes principiantes bien intencionados de que sus ideas son interesantes pero que (por ejemplo) preguntar a todos los adultos de una metrópolis de millones de habitantes qué piensan acerca de aumentar los impuestos en pro de la educación podría ser un poquito ambicioso. Es bueno hacer planes grandiosos pero, a menos que usted pueda traducir la pregunta a una dimensión manejable, es mejor que no comience. Si tales descomunales estudios iniciados por principiantes alguna vez llegan a realizarse (cosa que rara vez sucede, al menos en su forma original), casi siempre se convierten en una experiencia más negativa que positiva. A veces esos estudiantes se quedan con una montaña de datos pero sin material para una disertación. Aunque usted no esté buscando todavía un doctorado, de todos modos es una buena lección. Dése una oportunidad desde el principio y escoja una pregunta de investigación que sea razonable.

Por último, si hace algo que ya se hizo, es muy probable que pierda el tiempo. Existe una línea divisoria muy delgada entre lo que ya se llevó a cabo y lo que es importante que se haga con base en lo que ya se realizó. Parte del trabajo de investigación es aprender a construir sobre los resultados de investigaciones previas sin duplicar esas labores. Tal vez recuerde que al principio del capítulo hicimos hincapié en que la repetición es un componente importante del proceso científico y de las buenas prácticas de investigación. Su profesor puede guiarlo para

distinguir claramente entre lo que es redundante (hacer lo mismo sin una justificación sólida) y lo que es una contribución importante (hacer lo mismo pero explorando un aspecto de la investigación previa, o incluso hacer la misma pregunta pero eliminando fuentes de variación que estaban presentes en el primer estudio y que podrían haber confundido los resultados).

Defina sus intereses

Podría ser fácil para un investigador experimentado generar ideas para investigaciones adicionales, pero para eso le pagan (en parte, al menos). Además, los investigadores experimentados pueden aprovechar toda esa experiencia, y una cosa (un estudio) casi siempre conduce a otra (otro estudio).

Pero, ¿y el estudiante principiante como usted? ¿De dónde saca sus ideas para investigación? Aun si tiene el deseo ferviente de ser un psicólogo experimental, un profesor, un asesor o un trabajador social clínico, ¿dónde comienza a buscar sugerencias de ideas que tal vez querría investigar?

En algunos casos relativamente raros, los estudiantes ya saben desde el principio qué área de investigación van a escoger y qué preguntas de investigación desean hacer. Sin embargo, muchos otros estudiantes sienten más ansiedad y dudas que confianza. Antes de iniciar la importantísima revisión de la bibliografía, eche un vistazo a estas sugerencias respecto a dónde usted podría encontrar preguntas interesantes que valga la pena considerar como temas de investigación.

En primer lugar, las experiencias personales y los conocimientos de primera mano pueden ser el catalizador para iniciar una investigación. Por ejemplo, tal vez usted trabajó en un curso de verano con niños discapacitados y le interesa saber más acerca de la forma más efectiva de enseñar a estos niños. O bien, a través de sus propias lecturas a usted le ha nacido la curiosidad por el proceso de envejecimiento y cómo el proceso de aprendizaje cambia cuando una persona envejece. Al menos tres de mis colegas son educadores especiales porque tienen hermanos a los que no se les ofreció el servicio especial que necesitaban cuando niños para desarrollar todo su potencial. Sus propias experiencias moldean su personalidad. Sería una pena olvidarse del pasado personal al considerar el área general y el contenido de una pregunta de investigación, aunque no pueda percibir un vínculo inmediato entre esas experiencias y sus posibles actividades de investigación. Siga leyendo, y aprenderá cómo puede forjar ese vínculo.

Quizá usted desee asumir la responsabilidad plena de encontrar una pregunta de investigación. Por otra parte, nada tiene de malo acudir a su profesor o a otros miembros del personal académico de su escuela que estén trabajando en algún tema interesante y preguntarles: "¿Qué sigue?". Es probable que utilizar ideas de su asesor o profesor lo lleve a estar en la vanguardia de lo que está sucediendo en su campo. Ello también le ayudará a establecer y nutrir la importante relación entre usted y su asesor (o algún otro miembro del profesorado), necesaria para una experiencia fructífera y placentera. Éstas son las personas que están realizando las investigaciones, y mucho me sorprendería que no tuvieran más ideas que tiempo para ponerlas en práctica, y que no se alegraran de que un estudiante con inteligencia y energía (como usted) quiera ayudarles a extender sus actividades de investigación.

Otra posibilidad sería buscar una pregunta de investigación que refleje el siguiente paso del proceso de investigación. Tal vez ya se hayan hecho A, B y C, y lo que viene ahora es D. Por ejemplo, su interés especial podría ser entender los factores de estilo de vida que contribuyen a las enfermedades cardíacas, y ya sabe que factores como el tipo de personalidad (por ejemplo, Tipo A y Tipo B) y hábitos relacionados con la salud (por ejemplo, el consumo social de alcohol) han sido muy estudiados y sus efectos están bien documentados. El siguiente paso lógico podría ser examinar factores tales como hábitos de trabajo (incluidas la ocupación y la actitud hacia ella) o algún componente de la vida familiar (como las relaciones con el cónyuge). En las actividades de investigación de casi todas las disciplinas, y dentro de casi todos los temas, siempre existe ese siguiente paso que es preciso dar.

Por último, pero no por ello menos importante, está el hecho de que usted tal vez tenga que pensar en una pregunta de investigación porque se le pide en esta clase. Ahora bien, esto no es del todo malo, sino ve de esta manera: la gente a la que se le ocurren ideas fácilmente ya está del otro lado y no tiene que preocuparse por presentar una idea antes de una fecha límite. Las personas a las que les cuesta trabajo generar ideas necesitan una fecha límite, pues de lo contrario nunca harían nada. Así, pues, aunque hay razones más elevadas para pensar en preguntas de investigación, a veces hay que hacerlo simplemente porque se nos exige. Aun si éste es su caso, esfuércese por seleccionar un tema que pueda formular como pregunta de investigación y que retenga su interés durante toda la actividad.

Ideas, ideas, ideas (y qué hacer con ellas)

Incluso si usted está seguro de qué le interesa, a veces es difícil encontrar una idea específica para un proyecto de investigación. Para bien o para mal, usted es realmente el único que puede hacer esto, pero podría echar un vistazo a la siguiente lista de posibles temas de investigación. Para cada uno de estos temas existe un cúmulo de bibliografía asociada. Si un tema despierta su interés, consulte esa bibliografía (que se describirá en la segunda parte de este capítulo) y comience a leer. He aquí 61 temas; ¡seguramente uno de ellos hará que una campana suene en su mente!

abuso de drogas	nervioso	menarca
agresión	dietas	motivación
alucinaciones	diferencias de género	narcolepsia
aplicaciones de computadora	discapacidades del aprendizaje	nutrición
autoestima	divorcio	obesidad
biología de la memoria	dolor	optimismo
biorretroalimentación	edad adulta media	percepción
competitividad	educación bilingüe	pesadillas
condicionamiento clásico	egocentrismo	prejuicios
conflicto	epilepsia	programas de ventaja social
conjuntos mentales	ética	refuerzo
control de la natalidad	falsos recuerdos (déjà vu)	relajación
creatividad	identidad	ritmos circadianos
cuidado de los hijos	imagen corporal	SIDA
cuidado infantil	imágenes mentales	síndrome fetal de alcohol
cumplimiento	integración racial	sistema endocrino
depresión	inteligencia	sistema nervioso central
desarrollo cognoscitivo	inteligencia fluida	sueño con movimientos oculares rápidos (MOR)
desarrollo del dibujo	intervención temprana	sueños
desarrollo del lenguaje	mediación	visión a color
desarrollo del sistema	memoria	

De una idea a una pregunta de investigación y de ésta a una hipótesis

Una vez que usted haya determinado un posible interés específico, deberá apresurarse a formular una pregunta de investigación que desee estudiar, e iniciar su revisión de la bibliografía.

Existe una diferencia muy importante entre expresar un interés en una idea determinada y postular una pregunta de investigación. Las ideas están llenas de los productos de una actividad mental desbocada: creencias, preconcepciones, supuestos, imaginaciones, conjeturas y demás. Las preguntas de investigación son la articulación, de preferencia por escrito, de las ideas que al menos implican una relación entre variables. ¿Por qué es mejor hacer esto por escrito? Porque es muy fácil "salirse con la suya" al hablar. Es sólo cuando tenemos que poner en papel las cosas y atenemos a las consecuencias (las palabras habladas tienen la costumbre de desaparecer misteriosamente) que enfrentamos lo que hemos dicho, nos comprometemos y hacemos un esfuerzo por que lo escrito tenga lógica.

Una pregunta de investigación no es una expresión declarativa como una hipótesis, si una expresión de interés e intención claramente planteada.

Dado que todas las cosas se pagan tarde o temprano, cuanto más fácil de entender y más clara sea la pregunta de investigación, más fácil será el planteamiento de la hipótesis y la revisión de la bibliografía. ¿Por qué? Porque si desde el principio usted tiene una idea clara de lo que desea hacer podrá aprovechar su tiempo de manera mucho más eficiente cuando llegue el momento de buscar referencias y realizar otras actividades relacionadas con la revisión de la bibliografía.

Por último, llega el momento de formular una hipótesis o un conjunto de hipótesis que reflejen la pregunta de investigación. Recuerde que en el capítulo 2 presentamos una serie de cinco criterios que aplican a la postulación de cualquier hipótesis. Para refrescar su memoria, los volveremos a enumerar. **Una hipótesis bien escrita**

- se expresa en forma declarativa,
- postula una relación entre variables,

- refleja una teoría o un cuerpo bibliográfico en el que se basa,
- es breve y concisa, y se puede probar.

Cuando derive su hipótesis a partir de la pregunta de investigación, utilice estos criterios para probar si lo que usted está diciendo se puede comunicar fácilmente a otros y es comprensible.

Recuerde que las fuentes de ideas pueden ser cualquier cosa, desde un párrafo que leyó anoche en una novela hasta sus propios pensamientos únicos y creativos. Sin embargo, cuando usted llega a la etapa de la pregunta de investigación, es necesario adoptar una actitud más científica y expresar claramente en qué está interesado y cuáles variables considerará.

En la tabla 3.1 presentamos cinco intereses de investigación, las preguntas de investigación que se generaron a partir de esas ideas, y las hipótesis finales. Estas hipótesis sólo son finales en el sentido de que se ajustan más o menos a los cinco criterios arriba mencionados, pero es obvio que la revisión de la bibliografía y una discusión más detallada podría requerir una definición más precisa de las variables y tal vez hasta la inclusión de variables nuevas. Una buena hipótesis dice lo que vamos a hacer, no cómo lo vamos a hacer.

Revisión de la bibliografía

Y volvemos a lo mismo. Las investigaciones actuales se fundamentan en el trabajo arduo y la dedicación de investigadores del pasado y sus labores productivas. ¿Dónde encontramos los resultados de tales esfuerzos? En revistas y libros científicos, que encontramos (desde luego) en la biblioteca.

Si bien todas las etapas del proceso de investigación son importantes, una revisión lógica y sistemática de la bibliografía a menudo prepara el terreno para la creación de una propuesta de investigación bien hecha y para la realización de un estudio fructífero. ¿Recuerda uno de los errores fatales que mencionamos al principio del capítulo, el de escoger una pregunta de información que ya se hizo antes? ¿O el de seleccionar una pregunta trivial? Usted averigua todo esto y más cuando examina lo que ya se ha hecho y cómo se hizo. Una revisión completa establece un marco de referencia dentro del cual usted puede contestar la(s) pregunta(s) importante(s) que plantea. La revisión nos lleva cronológicamente a través del desarrollo de las ideas, muestra cómo algunas ideas se desecharon por falta de apoyo y cómo otras se confirmaron como verdaderas. Una revisión amplia y completa de la bibliografía nos proporciona esa crucial perspectiva para ver lo que se ha hecho y hacia dónde vamos, todo lo cual es indispensable para producir un informe bien escrito, bien documentado y bien planeado.

Interés o ideas de investigación

Problema o pregunta de investigación

Hipótesis

Aulas abiertas y éxito académico

¿Qué efecto comparativo tienen las aulas abiertas y las tradicionales sobre la lectura?

Los niños a los que se enseña a leer en situaciones de aula abierta leen en un nivel escolar más elevado que aquellos a los que se enseña a leer en una situación tradicional.

Habilidades para presentar exámenes, y calificaciones

¿Los estudiantes que saben cómo presentar un examen obtienen mejores calificaciones?

Los estudiantes que reciben capacitación en el método de "retención a corto plazo" obtienen mejores calificaciones en las pruebas de aptitud escolar que aquellos que no reciben tal capacitación.

Televisión y comportamiento de los consumidores

¿Cómo afecta el hecho de ver comerciales televisados la conducta de compras de los adolescentes?

Los muchachos adolescentes compran más productos anunciados en televisión que las muchachas adolescentes.

Abuso de drogas y abuso de niños

¿El abuso de drogas está relacionado con el abuso de

Existe una relación positiva entre el abuso de drogas entre los

	los niños?	adultos y el abuso físico y psicológico del que fueron víctimas cuando niños.
Cuidado de adultos	¿Cómo se han alistado muchos adultos a la responsabilidad de cuidar de sus padres ancianos?	El número de hijos que están cuidando de sus padres en el hogar del hilo ha aumentado en los últimos diez años.

Tabla 3.1 Ideas, preguntas e hipótesis. Nunca sabemos de dónde saldrán, así que lea sobre muchos temas y mantenga su mente abierta.

Así pues, prepare su cuadernillo de notas, sus tarjetas de referencia, un lápiz del número 2 y, tal vez, su computadora portátil, y comencemos. No olvide su credencial de estudiante para poder sacar libros de la biblioteca.

El proceso de revisión de la bibliografía consta de los pasos que se indican en la figura 3.2. Hay que comenzar con una idea lo más clara posible de qué se desea hacer, ya sea en la forma de un enunciado general acerca de las variables que se pretende estudiar, o como una hipótesis de investigación. El resultado final deberá ser un documento claro y bien escrito que justifique detalladamente por qué se escogió el tema en cuestión, cómo encaja con lo que ya se hizo antes, qué necesita hacerse en el futuro y su importancia relativa para la disciplina.

Básicamente, hay tres tipos de fuentes que usted consultará durante su revisión de la bibliografía (como puede ver en la tabla 3.2). Las primeras son las fuentes generales, que proporcionan pistas para localizar referencias de naturaleza general sobre un tema. Si bien es cierto que estas fuentes tienen limitaciones (que mencionaremos en breve), pueden ser valiosas porque ofrecen un panorama general y una introducción a un tema. Por ejemplo, digamos que a usted le interesa el área general de la psicología deportiva pero no tiene la

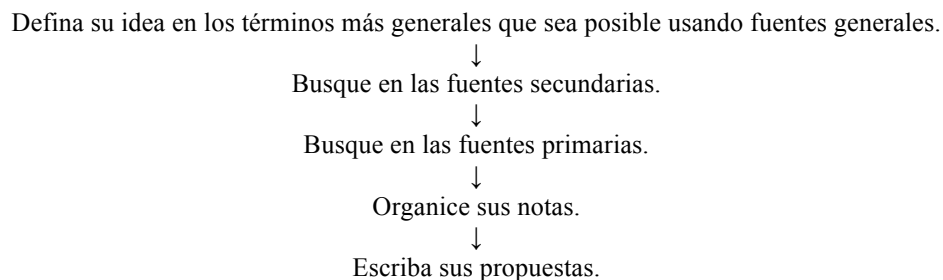


Figura 3.2 Pasos de la revisión de la bibliografía. Es una tarea imponente, pero si la desglosamos en sus partes veremos que está plenamente dentro de nuestra capacidad.

más remota idea de a dónde dirigirse para encontrar más información. Un punto de partida sería un artículo reciente que haya aparecido en un diario nacional y en el cual se mencione el nombre de uno de los psicólogos del deporte más afamados. Luego, usted podría acudir a fuentes secundarias o primarias más detalladas para averiguar más acerca del trabajo de esa persona.

El segundo tipo de fuentes comprende las fuentes secundarias. Estas fuentes están a «un paso» de distancia de las investigaciones propiamente dichas y son artículos de reseña, antologías de lecturas, síntesis de otros trabajos sobre el área, libros de texto y enciclopedias.

Las últimas y más importantes fuentes son las fuentes primarias. Éstas son informes de las investigaciones reales efectuadas, y aparecen como artículos de revistas científicas u otros tipos de obras originales, incluidos los resúmenes bibliográficos. En la tabla 3.2 se hace un resumen de las características de las fuentes generales, secundarias y primarias, y se dan ejemplos.

Fuente de información	Qué hace	Ejemplos
------------------------------	-----------------	-----------------

Fuentes generales	Proporciona un panorama sobre un tema dado y actúa como pista para encontrar información adicional.	Periódicos diarios, semanarios noticiosos, revistas populares, libros del ramo, Reader's Guide to Periodical Literature, New York Times index
Fuentes secundarias	Proporciona un nivel de información separado un paso de los trabajos originales.	Libros sobre temas específicos, reseñas de investigaciones
Fuente primarias	Los informes originales de los trabajos originales.	Revistas y libros científicos, resúmenes bibliográficos, ERIC

Tabla 3.2 Diferentes fuentes de información, qué hacen, y ejemplos. Las fuentes generales son especialmente valiosas como puntos de partida.

Empleo de las fuentes Generales

Las fuentes generales de información proporcionan dos cosas: una introducción general a áreas en las que usted podría estar interesado, y pistas respecto a los sitios donde se puede encontrar información más valiosa y útil (en un sentido científico, al menos) acerca del tema. También son un excelente material para hojear.

Cualquiera de las referencias que mencionaremos, sobre todo los índices de diarios nacionales, etc., pueden ofrecerle 5, 10 o 50 artículos sobre un área específica. En estos artículos casi siempre encontrará usted una buena introducción al área y una mención de algunas de las personas que están realizando investigaciones sobre el tema y dónde lo hacen. De ahí, usted puede examinar otros materiales de referencia para averiguar qué otras cosas han hecho esas personas o incluso ponerse en contacto directamente con ellas.

La biblioteca de su universidad seguramente tiene toneladas de fuentes generales, lo mismo que las bibliotecas públicas. He aquí una breve descripción de las fuentes que con más frecuencia se consultan, sobre todo en Estados Unidos, y una lista de otras fuentes que tal vez le interesaría consultar. Recuerde que debe usar las fuentes generales sólo para orientarse respecto a lo que existe y familiarizarse con el tema. Si bien los artículos periodísticos casi siempre son interesantes e informativos, y algunos están muy bien escritos, no pueden sustituir la lectura y el entendimiento de las investigaciones originales.

La *Reader's Guide to Periodical Literature* es, por mucho, la guía más exhaustiva que existe para la bibliografía general. Está organizada por tema y se publica mensualmente; abarca cientos de revistas especializadas (como el *New England Journal of Medicine*) y de interés general (como el *Scientific American*). Como los temas están ordenados alfabéticamente, es fácil encontrar fuentes de lectura sobre el tema escogido. En la figura 3.3 se muestra la traducción de parte de una página de *The Reader's Guide*. Como puede ver, en la página aparecen las entradas disponibles sobre el tema del embarazo adolescente. Observe que hay un encabezado general, Embarazo Adolescente, y debajo una lista de artículos específicos por título, y se indica dónde aparecieron.

Otra fuente general valiosa es *Facts on File* (FOF), que se publica en Nueva York desde 1941. FOF resume noticias que se publican en más de 50 diarios y revistas estadounidenses y de otros países, y es un magnífico lugar para averiguar si ha aparecido algo sobre su área de interés en estas publicaciones. FOF se publica semanalmente y su índice es acumulativo al año en curso, así que no deberá tomarle a usted más de unos minutos averiguar si hay información disponible.

EMBARAZO ADOLESCENTE

Véase también Madres adolescentes

La trampa del bebé [artículo de portada] E. Gleick y otros. il *People Weekly* v42 p38-55 O 24 '94

El mejor anticonceptivo [Programa de Prevención del Embarazo Adolescente de M. Carrera en Harlem] E. Gleick y otros, il por *People Weekly* v42p56 O 24 94

Tuve el bebé de quien me violó. *Sassy* v7 p35 N '94

Muñecas que parecen vivas y pueden hacerte llorar [muñeca adicta al *crack* Nena Piénsalo Bien] il *Newsweek* v124 p68 D 5 '94

El nombre del juego es vergüenza [es moralmente indebido que adolescentes no casadas

tengan hijos] J. Alter. il *Newsweek* v124 p41 D 12 '94
 Llamada de alerta [R. Jurmain, inventor de la muñeca Nena Piénsalo Bien]
 5. K. Reed, il por *People Weekly* v42 p103-4 O 10 '94
 Qué hacer cuando su hija adolescente le dice que está embarazada. K. A. Haynes. il *Ebony*
 v42 p 100+S '94

Figura 3.3 Entrada traducida de *The Reader's Guide to Periodical Literature*.

El *New York Times Index* se remonta a 1851 y lista todos los artículos publicados en ese diario por tema. Si usted encuentra una referencia a un artículo que podría interesarle (como se ve en la figura 3.4), puede ir a la hemeroteca y escoger una copia del número en cuestión o ver -claro, si está disponible el servicio- una versión microfilmada. Los originales casi nunca están disponibles porque se imprimen en papel delgado diseñado para soportar sólo los pocos días que la gente podría usar y prestar el diario.

Es por ello que el contenido se registra en microfilm o algún otro medio y se pone a disposición del público a través de las bibliotecas y hemerotecas. Muchas bibliotecas ofrecen lectores de microfilm que permiten copiar directamente de la imagen microfilmada y obtener una copia impresa de lo que se está consultando. Hoy día también puede obtenerse electrónicamente el texto completo de muchos diarios; hablaremos de esto un poco más adelante.

Nadie debe tomar lo que se imprime como la verdad absoluta, pero los semanarios noticiosos como *Time*, *Newsweek* y *U.S. News and World Report* ofrecen información general y mantienen al lector informado acerca de otros sucesos relacionados. Es posible que usted ni siquiera sepa que tiene un interés en un tema específico (como los aspectos éticos de la investigación). Un artículo sobre el tema podría aparecer en el número de esta semana, llamar su atención, y antes de que usted se dé cuenta estará utilizando esa información para buscar otras fuentes.

También hay algunas revistas especializadas que tal vez le convendría a usted conocer. *Science News* (que se publica semanalmente) y *Science Digest* (una publicación mensual) ofrecen resúmenes de noticias importantes del mundo de la ciencia. Estas publicaciones están al día y son informativas.

NIÑOS Y JÓVENES. Véase también

Accidentes y Seguridad, 5 20

Síndrome de Inmunodeficiencia Adquirida (SIDA), E 5,6,8,10,11,12, F 10,13,21,24,25, Mr 9,21,25, Ab 1,24,29, My 6,8, Jn 7,8,17,26,27, Ji 2,11, Ag 17,19, 5 26, 022,24, N 3,21, D 4

Adopciones

Publicidad, Ji 25

Afganistán, F 21,22

África, Jn 18,19

Agricultura, Ji 6

Aviones y líneas aéreas, Ab 6,28,29, My 1, Ji 13, 5 28, D 25

Abuso del alcohol, Jn 12

Bebidas alcohólicas, Mr 6, Ag 24, 5 28

Parques de diversiones, Ag 31, 5 1

Animales, Mr 1

Ropa, F 16, Ag 22,28

Arquitectura, E 20, My 5

Arte, E6, Mr23, Jn3, S4, 026,23

Asaltos, E 29,30, Mr 9, My 14, Jn 11, 5 8,30, O 24, N 30, D 14,15

Asma, E 4, My 8

Figura 3.4 Entrada traducida del *New York Times index*. Éste es el diaria más importante de Estados Unidos y generalmente cubre muchos avances importantes en todos los campos.

Por último, tenemos la mina de información que usted puede extraer de fuentes cotidianas como su diario local, boletines de las compañías y otras publicaciones. Los diarios locales a menudo publican los mismos artículos de la Associated Press que los principales diarios del mundo como el *New York Times* y el *Washington Post*. Tampoco olvide las oficinas de publicaciones del gobierno de su país, que publican con regularidad miles de documentos sobre todo tipo de temas, y que en buena parte son gratuitos. (No se preocupe; sus padres ya pagaron.)

Dentro de las publicaciones oficiales, las más útiles son los resúmenes estadísticos que contienen información de censos y otras fuentes. Generalmente estos resúmenes los publica el Ministerio de Comercio o su equivalente. Póngase en contacto con las oficinas de ese ministerio para averiguar qué publicaciones están disponibles.

Como usar las fuentes secundarias

Las fuentes secundarias son las que usted examina si quiere encontrar un resumen erudito de las investigaciones que se han realizado en un área específica o si busca fuentes de referencias adicionales.

Reseñas y síntesis de la bibliografía

En muchas bibliotecas existen unos libros GRANDES que suelen estar ordenados en la sección de referencia de la biblioteca (no en los anaqueles ordinarios). Dichos libros ofrecen reseñas y resúmenes de artículos u obras por área. He aquí un resumen de las más útiles.

The Review of Educational Research apareció por primera vez en 1973 y sigue siendo publicado por la American Educational Research Association trimestralmente. Esta publicación ofrece una colección de reseñas críticas de investigaciones en un área específica como los exámenes de competencia mínima (1983) o las cambiantes concepciones de la inteligencia (1983). Si bien es posible que esta publicación no ofrezca exactamente lo que usted busca, le dará una excelente perspectiva sobre la materia y usted da con el tema que le interesa.

También está The Encyclopedia of Educational Research, cuya versión más reciente se publicó en 1992 y que consiste en cuatro volúmenes que incluyen artículos escritos por expertos. Cada uno de estos artículos contiene una amplia bibliografía, así que es un buen lugar para comenzar a encontrar información sobre un tema en particular.

Una fuente secundaria general de reseñas bibliográficas es Annual Reviews of Psychology (publicada por Annual Reviews) que contiene unos 20 capítulos y abarca una amplia gama de temas. Imagínese; usted podría leer los últimos 10 años de estos volúmenes y estar perfectamente actualizado sobre una gama muy amplia de temas generales de psicología. Si encuentra un capítulo que trata exactamente lo que usted quiere hacer, ya habrá aventajado enormemente en su revisión.

Otra reseña anual que bien vale la pena considerar es la contenida en los anuarios de la National Society for the Study of Education (NSSE). Cada año desde 1902, esta sociedad dedicada al estudio de la educación ha publicado un anuario de dos volúmenes que se enfoca en un tema en particular, como adolescencia, microcomputadoras en el salón de clases, niños talentosos y superdotados o gestión de aulas. El área de enfoque suele ser algún tema contemporáneo, y si a usted le interesa ese tema, la información puede serle de enorme valor.

¿Está usted interesado en el desarrollo infantil? Busque el Handbook of Child Psychology (publicado en 1983 por John Wiley) de cuatro volúmenes. La cuarta edición de esta obra es el punto de partida para los estudiantes de psicología infantil y del desarrollo, estudiantes de educación de infantes, estudiantes de medicina y enfermería y otros dentro de un campo muy amplio. Los cuatro volúmenes individuales son

- Historia, teoría y métodos
- Psicobiología de la infancia y el desarrollo
- Desarrollo cognoscitivo
- Socialización, personalidad y desarrollo social

Un enfoque más amplio de las etapas de la vida es el que ofrece la serie de reseñas de *Life Span Developmental Psychology* (publicado por Academic Press). Estos diferentes volúmenes tratan temas tales como investigación y teoría, conocimiento y crisis vitales normativas.

Tampoco debe usted olvidar el gran número de libros eruditos que a veces tienen múltiples autores y han sido editados por un individuo o que fueron escritos en su totalidad por una persona (aunque en este último caso a veces se consideran como fuente primaria, dependiendo de su contenido). Utilice el confiable catálogo de tarjetas o fichas bibliográficas de su biblioteca (o su sistema de búsqueda computarizado, si lo tiene) para encontrar el título o autor que necesita.

Como usar las fuentes primarias

Las fuentes primarias son el alimento básico de la revisión de la bibliografía. Aunque usted puede obtener algunas ideas excelentes y una buena cantidad de información leyendo las fuentes secundarias, tiene que recurrir a las obras originales para obtener la información específica que necesita para hacer su propuesta y para que ésta sea aceptada.

De hecho, la mejor política es incluir principalmente fuentes primarias en su revisión de la bibliografía, con algunas fuentes secundarias para apoyar su propuesta, y que ni siquiera se le ocurra incluir fuentes generales. No es que la información de las revistas para el hogar o del diario local carezca totalmente de utilidad o valor; es de segunda mano, y no debemos basar nuestros argumentos en la interpretación de algún concepto que otra persona hizo.

Cómo usar las revistas científicas

¿Quiere usted revistas científicas? Nada más eche una mirada a la lista de la tabla 3.3 ordenada por categoría. Esto deberá ser suficiente para que usted pueda responder a su profesor cuando pregunte "¿Quién puede mencionar algunas de las revistas científicas más importantes en nuestro campo?" La lista es sólo una pequeña selección de lo que existe, usted deberá buscar otras más específicas o de más fácil consulta en su localidad.

Las revistas científicas son, por mucho, la fuente primaria más importante y valiosa de información acerca de un tema, ya que representan el vínculo más directo entre el investigador, el trabajo de otros investigadores y los intereses propios de usted.

¿Qué es realmente una revista científica y cómo funciona? Una revista científica es (casi siempre) una colección de artículos de investigación publicados en un área específica por algún grupo profesional. Por ejemplo, la American Psychological Association publica revistas que incluyen *The Journal of Experimental Psychology*, *Psychology*, *Public Policy and Law* y el *Journal of Counseling Psychology* (además de muchos otros). La Society for Research in Child Development publica *Child Development* y *Child Development Monographs*. Si usted se hace miembro de estos grupos recibirá las revistas como parte del paquete, pero también puede suscribirse exclusivamente a las revistas.

La mayor parte de las revistas científicas respetables funciona más o menos como sigue.

Primero, un investigador escribe un artículo ajustándose a un formato específico (como el que se muestra en el apéndice A) y luego envía tantas copias como exige el editor de la revista (por lo regular tres). Las pautas para preparar los manuscritos generalmente se encuentran al principio de cada número.

Segundo, una vez que el editor (quien suele ser un experto reconocido en el campo en cuestión) ha recibido el artículo, lo envía a por lo menos tres revisores que también son expertos en el campo. Estos revisores participan en un proceso llamado revisión ciega, ya que no saben quién es el autor (o autores) del artículo.

Tabla 3.3 de educación

Los nombres de los autores aparecen sólo en una hoja de portada que el editor arranca, y se usa el número de seguro social o algún otro número codificado para identificación. Esto hace que el proceso sea justo y que no haya la posibilidad de que las personalidades interfieran con lo que puede ser una meta altamente competida: publicar en las mejores revistas. Cada uno de los revisores hace una recomendación. Las opciones que tienen los revisores suelen ser parecidas a las siguientes:

- aceptación inmediata, lo que implica que se trata de un artículo sobresaliente y puede aceptarse para su publicación tal como está.
- aceptación con modificaciones, lo que significa que los autores deberán efectuar algunos cambios para que el artículo sea aceptado.
- rechazo con sugerencias de modificaciones, lo que significa que el artículo no es aceptable de momento, pero que se invita a los autores a volver a presentarlo después de haber hecho ciertos cambios, y
 - rechazo inmediato, lo que implica que el artículo es totalmente inaceptable.

Por último, si los revisores llegan a un consenso, el editor de la revista comunica esa decisión al autor o autores. Si no hay consenso, es tarea del editor tomar una decisión o enviar el artículo a otro revisor para recibir comentarios adicionales. Los editores trabajan intensamente para asegurar que los procesos de revisión y publicación en la revista sean justos. Los editores generalmente permanecen en su puesto de cuatro a seis años.

Por cierto, tal vez le interese saber que la tasa de rechazo promedio de las principales revistas científicas es de cerca de 80%. Si, 80% de los artículos que se presentan no logran salir publicados, pero los que son rechazados por las principales revistas casi siempre logran aparecer en otras publicaciones. El simple hecho de que un trabajo no haya sido aceptado por las revistas que tienen la tasa de rechazos más alta no implica que no se trate de informes útiles. De hecho, varios estudios han demostrado que hay muy poca consistencia entre los revisores, y que lo que podría recibir una alta calificación de uno podría recibir una muy baja de otro.

Una nota más acerca de las fuentes primarias en general. Si usted sabe de una revista científica o libro que podría necesitar, y su biblioteca no la tiene, no se desespere. Primero, visite otras bibliotecas cercanas o hable con algunos de los profesores de su departamento. Ellos podrían prestarle el libro o la revista. Si todo esto falla, es posible que su biblioteca participe en algún programa de *préstamos inter bibliotecarios*; el encargado de su biblioteca con gusto le dará informes al respecto. Este tipo de servicios ayudan a localizar y obtener físicamente de otra biblioteca los materiales de referencia deseados durante un tiempo limitado. Estos sistemas generalmente son rápidos y eficientes.

Uso de resúmenes bibliográficos

Sí las revistas científicas son la base de la revisión de la bibliografía, las colecciones de resúmenes (*abstracts*) no pueden quedarse muy lejos en cuanto a comodidad y utilidad. Un **resumen bibliográfico** es un resumen de un párrafo (o cuando más dos) de un artículo publicado en una revista científica; contiene toda la información que los lectores podrían necesitar para decidir si les conviene o no leer el artículo completo.

Un examen de las colecciones de resúmenes puede ahorrar a los investigadores una buena cantidad de tiempo, en comparación con hojear las revistas de las cuales se extraen los resúmenes. La mayor parte de las colecciones de resúmenes incluye también índices temáticos y por autor para ayudar a los lectores a encontrar lo que están buscando, y los resúmenes de artículos generalmente aparecen en más de una colección de resúmenes. Por ejemplo, un estudio sobre cómo tratar los niños que causan problemas en grupos podría aparecer en *Psychological Abstracts* (habiendo sido publicado en una revista como *Perceptual and Motor Skills*) y también en el *Current Index to Journals in Education* (por haber aparecido en una revista como *Psychology in the Schools*). No se preocupe si hay un traslape. De hecho, esto quiere decir que usted está realizando una revisión exhaustiva.

He aquí una breve descripción de algunas colecciones de resúmenes que podrían resultarle útiles.

El abuelito (¿abuelita?) de todas las colecciones de resúmenes es *Psychological Abstracts*, publicado por la American Psychological Association. *Psychological Abstracts* reseña y resume con regularidad más de 1000 (cuéntelas) revistas científicas en las 16 áreas siguientes:

- Psicología general
- Psicométrica
- Psicología humana experimental
- Psicología animal experimental y comparativa
- Psicología fisiológica
- Intervención fisiológica
- Sistemas de comunicaciones
- Psicología del desarrollo
- Procesos y problemas sociales
- Psicología social
- Desórdenes físicos y psicológicos
- Tratamiento y prevención

- Personal profesional y cuestiones profesionales
- Psicología de la educación
- Psicología aplicada
- Psicología del deporte y el esparcimiento

Hay tanta información en *Psychological Abstracts* que se hizo necesario crear un método más eficiente para localizar la información más allá de un sencillo índice de materias. Aquí es donde entra el *Thesaurus of Psychological Index Terms* (Tesauro de Términos de Índices Psicológicos), como se muestra en la figura 3.5. Antes de comenzar a usar *Psychological Abstracts*, usted usa el tesauro para seleccionar palabras clave para su búsqueda; luego acude a los resúmenes reales y busca bajo las palabras clave que identificó. Por ejemplo, si usted está interesado en el embarazo, encontraría las siguientes palabras clave (o términos relacionados) en el tesauro: nacimiento, adiestramiento para el parto, fertilización, complicaciones obstétricas, placenta y más.

Embarazo67
 PN2063
 UF Gestación
 N Embarazo adolescente88
 R Nacimiento64
 Adiestramiento para el parto78
 Fertilización73
 Complicaciones obstétricas78N
 Placenta73
 Periodo postnatal73
 Tecnología reproductiva88
 Reproducción sexual73

Figura 3.5 Entrada traducida del tesauro de *Psychological Abstracts*.

Otra forma de usar *Psychological Abstracts* es consultar la palabra clave *bibliography*. Bajo este encabezado usted encontrará una lista de obras y artículos que se han publicado; con suerte, encontrará la que se enfoca en su área de interés. Dos índices de especial utilidad (publicados por el Educational Resources Information Center o ERIC) son *Resources in Education* y *Current Index to Journals in Education*, cada uno con su propia función.

Resources in Education (RIE) presenta los resúmenes de artículos que se han presentado en conferencias, los resultados de investigaciones internas que incluyen informes de avance durante la realización de proyectos y otros documentos que no aparecen generalmente en las publicaciones formales. Muchos de estos trabajos finalmente se publican y luego entran en colecciones de resúmenes como *Psychological Abstracts*.

Current Index to Journals in Education (CIJE) no es realmente un índice sino un conjunto de resúmenes de más de 750 revistas científicas que se enfocan hacia el campo de la educación según una definición muy amplia. En este caso también, cabe esperar que estos resúmenes aparezcan en fuentes adicionales.

Al igual que *Psychological Abstracts*, el sistema ERIC opera con un conjunto de términos descriptivos que se encuentran en un tesauro, el *Thesaurus of ERIC Descriptors*, que siempre debe ser su primera opción antes de usar *RIE* o *CIJE*. Una vez que haya encontrado las palabras o los descriptores de su búsqueda, utilice el índice de materias (que se publica mensualmente) hasta encontrar el número de una referencia que le parezca interesante. Por último, consulte la descripción real de la referencia. Si usted quiere una copia del documento completo representado por el resumen, puede solicitar una copia impresa o microfilmada (más pequeña y económica) a través del Servicio de Reproducción de Documentos de ERIC utilizando formatos que tal vez pueda proporcionarle su biblioteca. Generalmente tarda varias semanas recibir el documento, pero es posible que usted no necesite ordenar una copia. Su biblioteca podría tener ya el documento en cuestión. Otra posibilidad sería ponerse en contacto con el autor original, cuyo nombre aparece en el resumen.

ERIC ha estado funcionando desde 1981 y tiene 19 centrales regionales en Estados Unidos que archivan, resumen y diseminan artículos y documentos sobre educación. La definición de educación es muy amplia, y cubre muchas disciplinas de las ciencias sociales y del comportamiento. Podemos darnos una idea de la amplitud del alcance de ERIC examinando la lista de áreas que cubren las 19 centrales mencionadas:

- desarrollo adulto, de carrera y vocacional
 - educación artística
 - orientación y servicios personales
 - administración educativa
- educación primaria y de la infancia temprana
 - niños con discapacidades y superdotados
 - educación superior
 - recursos de información

- educación media inferior
- idiomas y lingüística
- alfabetismo
- habilidades de lectura y comunicación
- educación rural y escuelas pequeñas
- educación en ciencias, matemáticas y ambiental
- estudios sociales/educación en ciencias sociales
- educación de maestros
- exámenes, mediciones y evaluación
- estudios Estados Unidos-Japón
- educación urbana

¿Cree usted que eso es suficiente para comenzar?

Psychological Abstracts y los conjuntos de resúmenes de ERIC son recursos importantes, pero hay otros un poco más especializados que también son muy útiles.

En la figura 3.6 se muestra la traducción de un resumen de un número reciente de Child Development Abstracts & Bibliography. Usted puede ver que contiene la referencia completa del artículo y un resumen en un párrafo del contenido del artículo que incluye:

- un enunciado introductorio acerca del contenido del artículo
- una descripción de los participantes por edad y otros factores de interés, y

-un resumen de los resultados.

Child Development Abstracts & Bibliography resume más de 300 revistas científicas y proporciona reseñas de libros acerca de niños y familias. La cobertura abarca seis áreas:

- biología, salud y medicina
- cognición, aprendizaje y percepción
- psicología social y estudios de personalidad
- educación
- psiquiatría y psicología clínica
- historia, teoría y metodología

Los títulos de otras colecciones de resúmenes, como Sociological Abstracts (Resúmenes sociológicos), Exceptional Child Education Resources (Recursos de educación para niños excepcionales), Research Related to Children (Investigaciones relacionadas con niños) y Dissertation Abstracts (Resúmenes de disertaciones) revelan la amplia variedad del material de referencia que está disponible.

356. Olson, Heather Carmichael; Sampson, Paul D.; Barr, Helen; Streissguth, Ann

P. & Brookstein, Fred L. (1992). **Exposición prenatal al alcohol y problemas escolares en la infancia tardía: Un estudio prospectivo longitudinal.** Development and Psychopathology, 4, 341-359.

Este estudio de seguimiento de 458 singletons evalúa el grado en que los problemas de conducta y aprovechamiento de los niños de 11 años en el salón de clases pueden predecirse por la exposición prenatal al alcohol a lo ancho de todo el espectro de uso por parte de la madre. Los análisis revelan una relación dosis-respuesta significativa y sutil entre la exposición prenatal al alcohol y el desempeño escolar de los niños una década después. Las juergas de la madre y el beber durante los primeros días del embarazo son particularmente importantes para el desempeño deficiente del niño en la escuela. Una amplia variedad de conductas problemáticas en el salón de clases, que incluyen problemas de atención, actividad, procesamiento de información y académicos, se relacionan con la exposición prenatal al alcohol.

Figura 3.6 Entrada traducida de Child Development Abstracts & Bibliography que reseña más de 300 revistas científicas distintas.

Cómo usar los índices

Las revistas científicas y resúmenes bibliográficos proporcionan la sustancia de un artículo, ponencia o informe. Si usted quiere un panorama breve de dónde podrían encontrarse las cosas, debe recurrir a un índice: un listado alfabético de entradas por tema, autor o ambas cosas.

Un buen punto de partida para cualquier revisión de la bibliografía es examinar los trabajos de personas que están en la misma situación que usted, estudiantes de maestría o doctorado. Primero está el Comprehensive Dissertation Index, publicado por la University of Michigan, que lista las disertaciones y tesis estadounidenses para las que existen resúmenes. Es probable que el Ministerio de Educación de su país o una o varias de las universidades más importantes de su país publiquen también índices de disertaciones y tesis similares.

Otros índices estadounidenses son American Doctoral Dissertations y Master's Abstracts: A Catalog of Selected Masters Theses on Microfilm, también publicados por la University of Michigan. Aunque no son tan actuales como Comprehensive Dissertation Index, American Doctoral Dissertations lista los títulos de las disertaciones doctorales por tema y año, obtenidos de los programas de graduación.

Master's Abstracts proporciona resúmenes de tesis de maestría.

Los ampliamente utilizados y populares Social Sciences Citation Index (SSCI) y Science Citation Index (SCI) operan de una forma interesante y creativa. SCI se dirige al terreno de medicina, agricultura y tecnología. SSCI se ocupa de los campos de las ciencias sociales y de la conducta. Digamos que usted lee un artículo que le parece muy pertinente para su propuesta de investigación y desea saber qué otras cosas ha hecho el autor. Una cosa que podría hacer es buscar por tema en los resúmenes como ya explicamos, pero es posible que lo que usted desea es encontrar otros artículos del mismo autor o sobre el mismo tema general. Las herramientas como SSCI le permiten enfocar su tema específico y acceder a casi toda la información disponible. Por ejemplo, ¿quiere usted saber quién ha mencionado el artículo clásico "Rasgos mentales y físicos de mil niños superdotados" (escrito por Louis Terman y publicado en 1925)? Busque Terman, L., en SSCI año por año, y encontrará tantas referencias que quizá no sepa qué hacer con ellas.

Por último, usted puede consultar el Bibliographical Index, una compilación de bibliografías que es el resultado de una búsqueda en cerca de 2600 publicaciones periódicas. Nada más piense en el tiempo que podría ahorrar si encuentra una bibliografía relativamente reciente sobre el tema que a usted le interesa.

Como usar la computadora para realizar búsquedas en la bibliografía

Imagine esto por favor: usted está en su departamento ya entrada la noche, y se da cuenta de que necesita una referencia más sobre el desarrollo de la autoestima en los adolescentes para completar su revisión bibliográfica. Usted está cansado; hay una lluvia intensa o está nevando, la biblioteca está a punto de cerrar y de todos modos es muy probable que no tengan lo que usted necesita.

¡Zum! Su servicio de búsqueda computarizada en línea llega al rescate. Usted mete el software (o lo llama de su disco duro), marca el número, se conecta con el servicio, localiza la base de datos y busca la referencia. En 20 segundos usted ha obtenido la referencia para leerla o imprimirla. ¿Es real esto? Puede apostar que sí, y es muy posible que su profesor ya utilice un servicio de este tipo. También es casi seguro que para cuando usted haya terminado sus estudios de licenciatura o posgrado, esté volando raudo de una base de datos a otra y de un proveedor de información a otro utilizando su computadora personal y hardware y software sencillo y de bajo costo.

En el hogar, en la oficina o en los confines de una biblioteca el uso de computadoras para llevar a cabo búsquedas y revisiones bibliográficas está floreciendo, y cada día hay más base de datos en las que se puede buscar.

Búsquedas fuera de línea con un CD-ROM

En muchas universidades usted puede entrar en la biblioteca principal, sentarse frente a una computadora de bajo costo, insertar uno de los discos CD-ROM (disco compacto-memoria sólo de lectura) que contienen referencias (como los de ERIC) y explorarlos en segundos en busca de la referencia que usted necesita. No está mal. También es posible acceder a discos CD-ROM que están almacenados centralmente.

Por ejemplo, he aquí lo que podría suceder durante una búsqueda de referencias sobre agotamiento de maestros de escuela primaria. Primero, se podrían teclear los caracteres maestros-agotamiento. El guión (o a veces un signo "+") indica que se deben buscar los registros relacionados con ambas palabras. En unos cuantos segundos aparece el número de registros hallados, digamos 366.

Luego, se teclearían los caracteres maestros-primaria, y aparecería el número de registros hallados, digamos 559.

Por último, para ver cuántos registros contienen ambos tipos de información, se podría teclear una instrucción (que en el caso de ERIC sería !1 2!) que le indique al sistema que debe buscar registros que contengan referencias tanto a agotamiento-maestros como a maestros-primaria. Digamos que en este caso se encuentran 10 referencias y se listan en la pantalla. Ahora usted puede imprimir esta información, leerla de la pantalla o guardarla en disco. ¿Fácil? Definitivamente. ¿Costoso? En algunos casos, es gratuito (léase: ¡se paga vía impuestos!) A medida que se haga necesario adquirir más discos y computadoras, es posible que se cobre por cada búsqueda.

Otra fuente increíblemente útil de información de referencia a la que usted puede acceder a través de su computadora personal y que está en un CD-ROM es Microsoft Bookshelf, un conjunto de obras de referencia que incluye una enciclopedia, un diccionario, una compilación de citas, un atlas, un almanaque, un tesoro y una tabla cronológica. En la figura 3.7 se muestra cómo se usó ese paquete para encontrar información sobre Sir Francis Galton, el famoso genetista. Si bien esta información no es suficiente para escribir el trabajo final, ciertamente basta para tener una idea de quién es esta persona e incluso para obtener información sobre qué hizo. Con un costo de \$77 dólares (para la versión Windows 95) es una ganga, e incluso es posible conseguirla en un paquete con otros productos Microsoft.

Búsquedas en línea

Muchas bibliotecas ofrecen recursos increíbles para acceder a información a través de computadoras personales ya sea aisladas (como en el caso de los discos de ERIC) o conectada a alguna otra computadora del campus o de otra ciudad.

Los investigadores de universidades, empresas y el gobierno están acudiendo cada vez más a proveedores de información en línea para

encontrar la información clave que necesitan, sea una referencia o dato específico, como el número de bicicletas producidas por Japón o el número de adultos jóvenes que viven en áreas urbanas.

El valor de las búsquedas en línea

¿Por qué realizar búsquedas en línea si puede obtener el mismo resultado recorriendo los pasillos de la biblioteca, revisando libros y revistas científicas, leyendo los resúmenes bibliográficos e índices (¿ya se cansó?) que mencionamos en la primera parte del capítulo?

Sin duda ya lo adivinó...el tiempo. Usted puede efectuar una búsqueda utilizando uno de los servicios que describiremos a continuación en la cuarta parte del tiempo que tardaría en hacerlo manualmente. Simplemente caminar hasta la biblioteca puede tomarle 30 minutos. Si maneja, gastará otros 15 minutos buscando un lugar para estacionar su automóvil. La lista es interminable.

Otra ventaja importante de las búsquedas en línea, si sus habilidades de búsqueda tienen cierto nivel de competencia, es que no es muy probable que omita mucho. Los proveedores de información (como CompuServe y America Online) ofrecen acceso a decenas de miles de documentos, ya sea en sus propias bases de datos o en otras a las que proporcionan acceso. Tenga presente también que ahora muchas universidades ofrecen acceso a sus bibliotecas desde fuera del campus, otra buena razón para adquirir habilidad en esta área. Por lo menos, podrá examinar un catálogo de existencias en línea. Si tiene suerte, también podrá acceder a dichas existencias.

Por último, y ésta podría ser la ventaja más atractiva, ésta es la ola del futuro. Hay tanta información en el mundo que pronto será casi imposible buscar de manera inteligente sin la ayuda de una computadora.

Sin embargo, el uso de los servicios en línea también tiene un aspecto negativo: el costo. Nada es gratis en este mundo, y esto incluye las búsquedas bibliográficas. Por ejemplo, en Estados Unidos, NewsNet, un servicio que proporciona texto en línea de cientos de diarios (antes de que salgan a la venta) cuesta ¡\$60 dólares la hora! Eso es mucho dinero para un individuo. CompuServe cobra \$2.95 dólares la hora (además de la cuota mensual de \$9.95 que incluye cinco horas gratuitas) y con práctica (para lo cual tienen un foro de práctica gratuito) usted puede entrar, obtener lo que necesita y salir rápidamente. Muchos proveedores de información tienen números telefónicos locales en las ciudades importantes, con lo que se minimizan los costos por uso del teléfono y sólo se paga por servicios adicionales como búsquedas especializadas. Si usted quiere jugar un juego de video por red contra un amigo que está en otra ciudad, diviértase; es probable que no le cobren extra. Pero si usted quiere realizar una búsqueda en Dissertation Abstracts, prepárese para pagar más. A medida que se familiarice con los sistemas, usted podrá buscar de forma más eficiente y ahorrará dinero.

Lo que se necesita para buscar en línea

Usted necesita cuatro cosas para iniciarse en el negocio de las telecomunicaciones:

- una computadora (de casi cualquier tipo)
- un módem
- software de comunicaciones
- una línea telefónica abierta

Lo primero es una computadora, que tal vez usted ya tenga. En cualquier actividad de telecomunicaciones la computadora actúa como transmisor y receptor de información. Asegúrese de que las características de su equipo sean las requeridas por su proveedor de servicios de búsqueda en línea.

Lo segundo es un módem. Un módem es un dispositivo que modula y desmodula señales electrónicas (que representan información) y transforma datos (sean el resumen de un artículo o un manuscrito entero) en señales que se pueden transmitir por una línea telefónica. El módem convierte la información digital en sonido, y se conecta entre la computadora y la línea telefónica. Ahí, el módem (y otro módem que está en el otro extremo de la línea telefónica) actúa como ágil intérprete y, cuando está funcionando bien, usted ni se da cuenta de que está trabajando. Por más avanzadas que sean las demás partes de su sistema de telecomunicaciones, si su módem es barato y no está a la altura de la tarea, el hardware más completo del mundo de nada servirá. Por ejemplo, algunos módems tienen filtros que eliminan parte del ruido electrónico que invade todas las líneas telefónicas. Cuando usted está hablando por teléfono, tal interferencia no es más que una molestia. Cuando usted se está telecomunicando, el ruido en la línea puede ser un desastre, porque echa a perder las transmisiones y lo que usted envía o bien no llega a su destino o llega tan alterado que ha perdido su significado. La calidad del software (que por lo regular viene junto con el módem) también es crítica, ya que algunos programas (como ProComm de DataStorm en Columbia, Missouri) hacen que el proceso de conexión sea rápido y sencillo, mientras que otros son tan frustrantes que no es sorprendente que algunas personas se den por vencidas y nunca vuelvan a intentarlo.

En tercer lugar, para que todo esto funcione, se necesita software de comunicaciones. Este es el software que traduce la información tal como usted la ve, a la forma en la que realmente se envía. Casi todos los módems vienen con algún software de comunicaciones, así como algunos sistemas operativos modernos como Windows 95. Además, es posible adquirir este tipo de software, el cual generalmente ofrece muchas más funciones que el que viene empacado con un módem.

Por último, usted necesita una conexión a través de una línea telefónica al sitio al que desea enviar (o del que desea recibir) información.

En otras palabras, necesita una línea telefónica. Ésta podría ser una línea dedicada exclusivamente a las telecomunicaciones o una línea normal que usted también usará para hablar. Si tiene sólo una línea tanto para voz como para datos, no podrá hablar por teléfono mientras está transmitiendo. La mayoría de la gente que toma en serio las telecomunicaciones acaba por contratar dos líneas telefónicas.

Una mirada a algunos servicios en línea

Mientras que algunas personas utilizan las telecomunicaciones para compartir información directamente con otras personas, aquí nos concentraremos en obtener acceso a proveedores de información, como CompuServe, América Online, Prodigy o la relativamente nueva Microsoft Network y Wow!. Estos son básicamente puertas de entrada a foros o grupos de discusión privados y enormes bases de datos en líneas en las que usted entra a través de su hardware y software de telecomunicaciones, o el software que el proveedor de información distribuye. Usted también puede aprovechar una tonelada de servicios distintos que ofrecen estos proveedores, como correo electrónico, grupos de noticias, foros y más (de todo lo cual hablaremos más adelante en este capítulo).

He aquí una breve descripción de los servicios más grandes y de lo que ofrecen. Casi todos ofrecen una inscripción gratuita y algunos tienen números 800 (de larga distancia sin costo para quien llama) para solicitar un juego de arranque.

CompuServe (800/848-8990 en Estados Unidos) le permite entrar de lleno en las actividades en línea con 10 horas de tiempo libre cuando usted se inscribe. Con más de 4 millones de suscriptores, CompuServe ofrece noticias, informes del tiempo, deportes, viajes, compras, correo electrónico, información financiera y acceso a otros servicios de información que incluyen más de 900 bases de datos bibliográficas (¿necesita usted encontrar ese artículo?). CompuServe es, por mucho, el servicio más completo orientado a la investigación. En la figura 3.8 usted puede ver una muestra de su pantalla inicial y de lo que está disponible.

Si usted es un usuario de Macintosh, CompuServe le ofrece Navigator, una herramienta de software que incrementa considerablemente la velocidad y la potencia de las actividades de búsqueda. Usted puede seleccionar su destino (por ejemplo, la base de datos que desea consultar y las palabras clave que quiere buscar) antes de ponerse en línea, y luego ejecutar ese guión a través de Navigator para llegar a su destino por un costo mucho menor. Para los usuarios de DOS y Windows, CompuServe ofrece el CompuServe Information Manager, una herramienta similar que le permite realizar las mismas cosas. Al igual que la mayor parte de los servicios, CompuServe ofrece búsquedas en bases de datos por una cuota adicional.

A medida que se popularizan los servicios de información, la gente innovadora está diseñando software que ayuda a los usuarios a pasar menos tiempo en línea y ser más eficientes. En el caso de CompuServe se cuenta con TAPCIS (una utilidad de DOS que puede obtenerse de Support Group, Inc., al 800/872-4768). TAPCIS le permite crear y leer correo fuera de línea (sin costo) y también configurar sus búsquedas antes de ponerse en línea, con lo cual se pueden lograr ahorros sustanciales.

América Online (AOL, 800/827-3338), el servicio en línea más popular y el que cuenta con más suscriptores, tiene más de 6 millones de miembros y es el lugar obligado al que usted busca es acceso rápido a una pléyade de diarios (como The New York Times) además de un gran número de bases de datos (aunque no tantas como con CompuServe). AOL es conocida por sus foros especiales que se ocupan de entretenimiento y deportes. Una mayor proporción de la energía invertida en el diseño y ejecución de AOL está dedicada a las necesidades de los consumidores generales, y no a las del investigador. No obstante, sí es posible acceder a bases de datos y obtener información, y es posible que AOL sea la fuente general definitiva.

Por último, está la relativamente nueva Microsoft Network (800/426-9400). Microsoft ha limitado su base de suscriptores a 500000 temporalmente. Como usted puede ver en la figura 3.9, este servicio ofrece noticias de actualidad además de abundantes conexiones para obtener información sobre temas que van desde las artes hasta los negocios y finanzas.

Cualquiera de estos proveedores de información en línea contiene más material del que usted jamás podría examinar y probablemente una buena cantidad de información relacionada con su interés exacto. ¿Cuál es el mejor para usted? Yo diría que CompuServe es el mejor en términos de volumen de información y acceso a diferentes bases de datos. Los otros dos ciertamente tienen mucho que ofrecer. Si a usted le interesa, tal vez lo mejor sea suscribirse a cada uno durante su oferta de prueba de 10 horas. Trate de encontrar a través de cada servicio información acerca de un tema que le interese, y vea cuál es más completo. Sólo asegúrese de que sea fácil cancelar la suscripción, ya que los proveedores del servicio le pedirán una tarjeta de crédito (que se usa para facturar) en el momento de inscribirse. Por ejemplo, la única forma de cancelar una suscripción a AOL es llamando a su número 800. ¡Ni siquiera es posible cancelar la suscripción en línea!

Como usar Internet para investigación

Usted ha oído hablar de ella e incluso tal vez la conoce, pero de todos modos permítanos presentarle a: ¡la Internet! Indescribiblemente misteriosa, ha recibido todo tipo de publicidad, tanto correcta como exagerada, y también ha recibido duras críticas. ¿Internet es para usted? Creo que puedo demostrarle que la respuesta es "sí". He aquí por qué.

¿Qué es Internet?

En los términos más básicos, la Internet (a la que también se le llama comúnmente la Red o simplemente Internet) es una red de redes. ¿Qué es una red? Una colección de computadoras que están conectadas entre sí y que pueden comunicarse unas con otras.

Imagine que todas esas redes estuvieran conectadas entre ellas. Imagine cientos de redes y miles de computadoras de todos tipos comunicadas unas con otras, y millones de personas usando esas computadoras. Ahora tiene usted una idea del tamaño de Internet. De hecho, Internet crece en más de 10% cada mes, y ya cuenta con unas 30000 conexiones (cada una de las cuales podría ser utilizada por cientos de personas) activas.

Actividades de investigación en la red

Si estamos hablando de información de todos colores y sabores, no hay mucho que no se pueda hacer en la Red. He aquí un breve panorama de cómo puede usarse la Red para fines de investigación, sea para ponerse en contacto por correo electrónico con un experto en su campo o para encontrar la página base (hablaremos más de las páginas base en breve) de la universidad a la que usted le gustaría entrar para sus estudios de posgrado, o para obtener una bibliografía sobre un tema en particular.

- El uso más frecuente de Internet probablemente es el correo electrónico o e-mail. Así como usted intercambia cartas con sus colegas al otro lado del país o en otros lugares del mundo, puede hacer lo mismo sin jamás tocar pluma ni papel. Usted crea un mensaje y lo envía a la dirección electrónica de su corresponsal. El método es rápido, fácil y divertido. Por ejemplo, si usted quiere un reimpreso de un artículo que le parece interesante, puede solicitarlo por correo electrónico al autor. Son cada vez más los miembros del personal académico y administrativo de instituciones educativas tanto grandes como pequeñas que tienen correo electrónico.
- Información, información, información: toda disponible en millones de archivos en la Red. Por medio del protocolo de transferencia de archivos o ftp, usted puede descargar o transferir de otras computadoras a la suya, archivos enteros. Con un poco de práctica, usted navegará en la Red para obtener todo, desde la colección de historia de la University of Kansas hasta la letra de los éxitos de rock and roll de los años sesenta. Por ejemplo, muchos investigadores colocan sus datos en un archivo que después puede transferirse fácilmente de la computadora del investigador a la computadora de cualquier otra persona, con objeto de que el acceso a los datos sea rápido y fácil.
- Existen miles de grupos de noticias electrónicos a los que usted puede suscribirse en Internet. Éstos son lugares donde se puede publicar y compartir información entre usuarios de Internet, con temas que van desde la exploración del espacio hasta la autenticidad de las escrituras de propiedad de un terreno que datan de la Guerra Civil estadounidense. Usted puede visitar y contribuir con todos estos grupos de noticias. Por ejemplo, si a usted le interesa el programa de estudios de matemáticas K-12, haga la prueba con los grupos de noticias k12.ed.math. ¿Le interesa la conducta patológica? Pruebe el grupo de noticias sci.psychology.psychotherapy.
- Cuando usted usa telnet para conectarse con un sitio remoto, es como si estuviera usando una computadora en ese otro sitio. Con una conexión de Internet, usted puede controlar computadoras a miles de kilómetros de distancia y obtener la información a la que tienen acceso esas computadoras. Por ejemplo, en una ocasión que yo quería saber qué otros trabajos había llevado a cabo cierto autor me conecté por telnet con la Biblioteca del Congreso de Estados Unidos e introduje el nombre del autor. En un segundo, había obtenido una lista de sus otros 40 títulos.
- Por último, está la World Wide Web o WWW, de la que todo mundo ha estado hablando. Aquí usted puede usar un navegador (browser) para establecer una conexión con estas paradas gráficas en la supercarretera de la información. Usted puede acceder a la página base de, digamos, una institución médica y ver qué tipos de programas de financiamiento están disponibles, o visitar el sitio de una universidad para averiguar los horarios de las materias que se están impartiendo y quién las imparte.

En este capítulo limitaremos nuestros comentarios al correo electrónico, los grupos de noticias y una exploración de la World Wide Web. También presentaremos ejemplos de como usar estas funciones de Internet. Es probable que el sistema para conectarse con internet y usarla en su escuela o en alguna institución que ofrezca el servicio, sea diferente de lo que describiremos aquí, pero los principios son exactamente los mismos.

Correo electrónico para todo mundo

Imagine que es 1925 y usted está sentado ante su escritorio en la universidad escribiendo una carta a un amigo en Europa. Usted pega una estampilla a la carta, la pone en el correo y unas semanas más tarde le llega una respuesta. Usted se sorprende de lo rápido que es el correo y se sienta para responder a las preguntas que le hace su amigo acerca de cómo le está yendo en la universidad y qué va a hacer después de graduarse.

Ahora imagine que es 1998 y usted le está escribiendo a un amigo en Europa, sólo que esta vez usted usa correo electrónico o e-mail. Desde su casa, usted redacta el mensaje, oprime la tecla de enviar y su amigo lo tiene en su escritorio en unos cuantos minutos. El mensaje de respuesta llega en otros 20 minutos más, y usted no cabe en sí de asombro. En este capítulo mostraremos lo fácil que es enviar correo utilizando el programa Eudora, que cuenta con versiones para Windows y para Mac.

El correo electrónico funciona de forma muy parecida al correo convencional. Usted es-cribe un mensaje y lo envía a una dirección. La gran diferencia es que no se usa papel. Más bien, los mensajes que usted envía viajan de una computadora a otra en cuestión de minutos u horas, en lugar de días o semanas, con la misma rapidez con que su voz viaja en una conversación telefónica.

Eudora y correo electrónico

Se requiere un programa de correo electrónico con todas las de la ley para gestionar realmente todas las actividades de correo.

Presentaré un ejemplo de sesión con Eudora en la que envío un mensaje a un investigador que yo sé que está trabajando en mi área de interés, la diseminación de información pública. He aquí una demostración en la que le escribo a un colega y le solicito una reimpresión.

Como puede ver en la figura 3.10, he redactado un mensaje a un colega para solicitarle una copia de un artículo. La pantalla de Eudora es parecida a la de muchos otros programas de correo en la que se redacta un mensaje.

- tiene un lugar para la dirección de Internet a la que se está enviando el mensaje (Judy Turner).
- Muestra de quién proviene el mensaje (njs@alcon.cc.ukans.edu, que en la vida real es Neil J. Salkind).
- Indica el tema o asunto del mensaje (Copia de artículo).
- Tiene espacio para enviar copias del mensaje a otras personas (Cc:)
- Tiene espacio para una copia ciega (Bcc:)
- Tiene espacio para un anexo (como un archivo o algún otro documento).

El contenido del mensaje aparece en el área de mensaje principal y comienza con "Estimada Dra. Turner". Una vez que el mensaje está completo, basta con hacer clic en el botón «Send» que está en la parte de arriba de la pantalla para enviar el mensaje a Judy Turner

Por lo regular; cuando se envía cualquier tipo de correo, sea electrónico o "tortuga" el destinatario responde. Eudora examina automáticamente mi buzón cada 30 minutos y me avisa si tengo correo que no he leído. Si así es, suena una campanita y puedo seleccionar la opción de Eudora que me permite ver el correo que ha llegado como se muestra en la figura 3.11.

Estimada Dra. Turner:

¿Seffa tan amable de enviarme una copia del artículo más reciente que usted publicó sobre diseminación de la información, 'en el Journal of Information Services? Agradezco su inapreciable ayuda.

Neil Salkid
213 Bailey Hall
University of Kansas
lawrence, Ks 66045

Como se observa en la figura, tengo un mensaje de Judy Turner que espera ser leído (eso es lo que significa el punto (.) que está a la izquierda del mensaje).

Para leer el correo, hago doble clic en el mensaje y puedo ver su contenido como se muestra en la figura 3.12.

En cualquiera de las pantallas de Eudora que se muestran puede verse que hay muchos botones en los que se puede hacer clic y un buen número de menús con muchas opciones para escoger. No voy a describir todas éstas, ya que mi propósito no es enseñarle a usar Eudora sino sólo mostrarle cómo funciona el correo electrónico. No obstante, es conveniente que usted tenga alguna idea de lo que un buen programa de correo electrónico puede hacer y Eudora es uno de los mejores.

Las funciones de Eudora le permiten

- guardar e imprimir mensajes de correo.
- reenviar correo a una o hasta cien personas más.
- incluir su "firma" en todos los mensajes que envía. Ésta puede ser un membrete, una frase célebre o su dirección de correo electrónico o lo que a usted le parezca mejor. Por ejemplo, he aquí la firma de uno de mis estudiantes, la cual aparece automáticamente al final de todos los mensajes que envía.

```
*
*Jim Williams      phone:(817)555-4532      *
* Summit High School  voice mail: (817) 255-4399 ext. 52      *
*2318Bluepkwy      i fax:(817)555-3419      *
* Webster, MO 64063      *
* www: http://www.place.k12.mo.usl      *
```

- tener tantos buzones como desee.
- revisar la ortografía.
- usar seudónimos (en vez de «jt@kuhub.cc.ukans.edu» puedo usar "Judy Turner")
- almacenar correo para leerlo después.

Una introducción a los grupos de noticias

Imagine poder encontrar información acerca de unos 30000 temas que van desde sistemas estereofónicos a chistes (censurados o no), ética legal, fútbol americano universitario o astronomía. ¿En qué parte del mundo podría usted encontrar una colección de información tan diversa y de fácil acceso? Adivinó. En Internet y en los diversos sitios de USENET que envían noticias todos los días a todos los rincones del mundo. Las noticias que se ajustan a una categoría, como fútbol americano universitario o ética legal, forman un grupo de noticias (también llamado grupo). Un grupo de noticias no es más que una colección de información acerca de un tema.

Para ayudar a controlar el flujo de artículos, los sitios de noticias son gestionados, moderados, administrados y censurados por administradores de sistemas que trabajan para instituciones como universidades y corporaciones. No todos los grupos de noticias llegan a todos los posibles sitios o a todos los usuarios que tienen acceso a un sitio Internet. Los grupos de noticias de los cuales usted puede seleccionar noticias son aquellos que el administrador del sistema pone a su disposición.

¿Qué hay en las noticias?

Los grupos de noticias se nombran y organizan de acuerdo con un conjunto de reglas. La más general de estas reglas tiene que ver con el nombre del grupo mismo. Los nombres de los grupos de noticias tienen una estructura jerárquica, donde el nivel más alto de la jerarquía aparece en la posición más a la izquierda. Por ejemplo, el nombre del grupo de noticias k12.ed.tech significa que dentro de k12 (el nombre general para el grupo de noticias que abarca desde el jardín de niños -kindergarten- hasta el duodécimo año -equivalente al último año de bachillerato-) hay un subconjunto llamado ed (de educación) y dentro de ese otro subconjunto llamado tech (de tecnología).

Nombre del grupo	Orientación general	Ejemplos de grupos
comp	Información sobre computadoras, informática, software y temas de computación de interés general.	- comp.champ (inteligencia artificial) - comp.lang.c (el lenguaje de programación C) - comp.bbs.misc (tableros de anuncios)
news	Información acerca de noticias, grupos de noticias y (a red de grupos de noticias.	news.adm (noticias de admon.) - news.ao.members (miembros de cierto grupo de noticias)
rec	Información sobre recreación, msatiempos, las artes interpretativas y otras cosas divertidas.	rec.sports.footba)) (para aficionados a fútbol) americano rec.arts.movies (reseñas y discusión sobre cine) - rec.audio (audio de alta fidelidad)
sci	Información sobre ciencia, investigación y descubrimientos científicos, ingeniería y algo de ciencias sociales.	sci.astro (astronomía) sci.med (medicina) sci.skeptic)OVNIs y otras especulaciones)
soc	Información sobre las ciencias sociales.	soc.history (historia) - soc.roots (genealogía) - soc.women (temas relacionados con la mujer)
talk	Foros sobre temas controvertidos y cuestiones para debate.	- talk.abortion (aborto) - talk.rumors (control de rumores) - talk.religion (temas religiosos)
misc	Todo lo que no entra claramente en alguna de las categorías anteriores.	- misc.taxes (impuestos) - misc.jobs (empleo) - misc.fitness (condicionamiento físico)

Tabla 3.4 Nombres de los grupos de noticias principales, sus áreas de enfoque generales y ejemplos.

Existen siete grupos de noticias principales, cada uno de 105 cuales utiliza un rótulo específico para el primer nivel de la jerarquía. En la tabla 3.4 usted puede ver los nombres de esos grupos y ejemplos de lo que se incluye en cada categoría.

Lo que se muestra en la tabla 3.4 es una lista de los grupos de noticias principales. También existe un conjunto alternativo de grupos de noticias, algunos de los cuales son aún más populares que los principales. Estos grupos alternativos, que se muestran en la tabla 3.5, son muy activos y también pueden proporcionar información valiosa.

Ahora usted ya sabe qué esperar con base en el rótulo que está al principio del nombre de un grupo de noticias.

Para ver cómo funciona un grupo de noticias, sigamos un ejemplo de alguien que se interesa en la tecnología de la educación y que utiliza

NewsXpress, un lector de noticias muy popular basado en Windows.

<i>Grupo de noticias alternativo</i>	<i>Orientación general</i>
alt	Cualquier cosa que se salga de lo común
bionet	biología
bit	La red de correo BITNET
biz	negocios
clari	Categoría específica no cubierta por otros grupos de noticias
de	Información en alemán
fj	Información en japonés
gnu	Proyecto operado por la free software foundation
ieee	Ingeniería eléctrica
k12	Temas de interés desde kindergarten hasta 12 años
relcom	Idioma ruso
umsnet	El sistema de computo VMS

Tabla 3.5 Nombres de los grupos de noticias alternativos, su orientación general y ejemplos.

Cómo usar NewsXpress para obtener noticias

NewsXpress es un lector de noticias que se ofrece como shareware y que cuenta con un juego muy completo de herramientas para leer y participar fácilmente en los grupos de noticias. Las herramientas también ayudan a ver todos esos miles de grupos de noticias para que usted pueda escoger los que le interesan.

En la figura 3.13, que muestra la ventana inicial de NewsXpress, pueden verse los grupos de noticias a los que estoy suscrito. En NewsXpress y otros lectores de grupos de noticias por lo regular se pueden ver todos los grupos a los que el administrador del sitio del usuario permite el acceso. Entre ellos, el usuario puede escoger aquellos a los que se desea suscribir. Cada vez que usted inicia su lector de noticias obtiene la versión actualizada del grupo, lo que incluye todas las noticias que se han agregado al grupo desde la última vez que usted lo abrió. Las versiones más recientes de los navegadores de la World Wide Web, como Netscape, también permiten acceder y participar en grupos de noticias.

El siguiente paso sería abrir el grupo de noticias k12.ed.tech y examinar su contenido, como se aprecia en la figura 3.14. Dentro de cada grupo de noticias aparece una lista de temas (en este caso 12), cada uno iniciado por un individuo como fuente de información adicional, lugar para reunirse electrónicamente, comentar cuestiones de interés, y demás.

Si alguien quiere participar en un grupo de noticias dado, puede añadir un nuevo tema en este nivel, o entrar en un grupo de noticias existente y hacer su propia contribución.

En la figura 3.15 puede verse el contenido de ANNOUNCE: Electronic Learning Magazine, que muestra la primera de nueve contribuciones que se han hecho. También en este nivel una persona puede responder a la contribución publicando un mensaje que se unirá a los otros nueve o escribiendo directamente al autor del mensaje (en cuyo caso la respuesta no se publicaría).

Cómo usar las listas de correo

Existe otra forma fantástica de utilizar los grupos de noticias, y es una magnífica fuente de información. Usted puede suscribirse a un grupo de discusión listserv, que es un depósito automático de información. Si usted se suscribe, recibirá todo lo que la lista recibe. Este servicio también se conoce como lista de correo listserv.

Por ejemplo, si usted pertenece a la lista de correo de tecnología de la educación desde kindergarten hasta el 12o. año, cada vez que alguien envíe correo a esa lista usted también lo recibirá. Hay más servidores de correo de los que usted puede imaginar, y se requiere algo de exploración para encontrar los que se ajustan a sus necesidades.

Para suscribirse a una lista de correo, es necesario enviar un mensaje al administrador de la lista. Apenas lo haga, comenzará a llegarle un flujo constante de mensajes. Cabe hacer una advertencia: si una lista es muy activa usted podría recibir cientos de mensajes en un día. Si pasa un día sin que usted examine su correo, es posible que su buzón se llene tanto que usted no pueda leer sus mensajes. Imagine su buzón físico afuera de su departamento o casa. Cuando se llena hasta el tope es muy difícil sacar cualquier cosa porque el correo está demasiado apretado. Se necesita una caja más grande, o vaciar la caja antes de que se llene demasiado. Lo mismo sucede con una lista de correo de Internet. Consígase un buzón de correo electrónico más grande (pida más espacio de almacenamiento al administrador de su sistema) o bien revise su correo más de una vez al día.

Introducción a la World Wide Web y Netscape

Ahora estamos listos para explorar realmente Internet y utilizar lo que para mucha gente es el aspecto más atractivo de Internet: la World Wide Web.

Usted ya sabe que la Internet es una red de redes. La World Wide Web (o WWW) es una colección de documentos que representan diferentes sitios enlazados entre sí, de modo que basta hacer clic en una palabra o imagen o sonido en particular para trasladarse rápidamente a otro sitio. En la Web usted encuentra los llamados documentos de hipertexto distribuidos, también conocidos como páginas base, que contienen vínculos. Estos vínculos conectan una página base a otra. Para ver y usar estas páginas base y vínculos se necesita un programa visualizador, y ahí es donde entra Netscape. Netscape es un navegador que puede servir para explorar la WWW. Existen otros visualizadores, como Mosaic y Cello, que funcionan bastante bien. La razón por la que escogí Netscape para este libro es que puede obtenerse en línea, lo que significa que usted puede probarlo sin costo antes de pagar y registrarse. Netscape está ampliamente disponible, y más gente lo usa que cualquier otro tipo de visualizador. Pero la mejor razón para escoger Netscape es que es muy fácil de usar y más divertido de lo que usted se imagina. Así pues, prepárese para navegar y conocer la World Wide Web.

Todo acerca de las páginas base

Una página base (home page, término traducido a veces como página hogar) es una colección de información con ciertas características comunes a todas las páginas que la conforman. Se puede llegar a las diferentes páginas base de diversas maneras, que se explicarán más a fondo en este capítulo. Por ahora, exploremos una de las páginas base iniciales de la Biblioteca del Congreso de Estados Unidos que se muestra en la figura 3.16. Esta página base tiene que ver con las exposiciones de la Biblioteca del Congreso.

En primer lugar, en la parte superior de la página vemos el título "Library of Congress World Wide Web Home Page". El título nos dice de qué trata la página base actual. El cuadro de texto Location (Dirección) muestra la dirección de la página base activa. Ésta es una dirección en la WWW y también se llama URL (universal resource locator) o localizador universal de recursos. Si usted conoce el URL de una página base dada, puede introducirlo en el cuadro de texto Location (Dirección), oprimir Enter, y Netscape lo llevará a esa página. Por su naturaleza, los URL son un tanto difíciles de entender, y no queda claro dónde está situada físicamente una página o qué institución la está patrocinando. Esto implica que resulta cómodo mantener una lista de los URL que a usted le gustan y desea visitar otra vez. Eso es exactamente lo que hace el menú desplegable Bookmarks (Favoritos).

Luego, la parte principal de la pantalla muestra el contenido de la página base, que inicia con una bonita imagen del edificio principal de la Biblioteca del Congreso. Como también puede verse en la figura 3.16, esta página base contiene, además del gráfico, palabras subrayadas que son vínculos con otras páginas (como la de qué hay de nuevo [what's new] o la de buscar [search]). Si usted hace clic en una de estas palabras resaltadas, Netscape lo lleva a una nueva página.

En la figura 3.17 se muestra una página base de Yahoo!, uno de los más útiles sitios de Web. Esta página casi no contiene otra cosa que vínculos con otras páginas base. Si usted hace clic en cualquiera de esos vínculos, se conectará con otra página base, que tiene más vínculos. La Web es interminable.

Dónde encontrar páginas base

Ésta es la pregunta de los \$64,000. No existe una lista central de páginas base, así que usted no puede recurrir a un directorio o alguna otra fuente y encontrar algo como «Todas las Páginas Base de la World Wide Web». Esto es imposible porque la Web y el número de páginas que contiene están cambiando rápidamente. Muchos libros ofrecen diferentes listas, pero algunos de esos libros se vuelven obsoletos antes de publicarse. Pese a ello, un libro así podría ser un buen lugar para comenzar.

La mejor manera de encontrar páginas base es explorar la Web. Hable con sus colegas para averiguar qué han descubierto, y utilice algún tipo de herramienta de búsqueda. Si usted encuentra una página base fabulosa, guarde su ubicación como marcador de página (bookmark) (favorito) y comparta esa información con sus amigos. Ahora bien, si usted dispone de tiempo, hay una página base muy buena, llamada URouLette, que le recomiendo explorar. Esta página base, creada en la University of Kansas, le permite hacer clic en una rueda de ruleta (el nombre de la página es un juego de palabras entre URL y roulette) e ir a una página base. ¿Cuál? La selección es al azar, así que usted podría encontrarse en una página base sobre rayos X de Uganda o sobre física solar y estelar de Harvard (<http://CFA-www.harvard.edu/CFA/ssp.html>), o en la de la miniescuela canadiense Point Gray Mini School de Vancouver, Columbia Británica (<http://trinculo.educ.sfu.ca/pgrn/pgmhome.html>), o en "Joyas de los Escritos de Baha'u'llah" (<http://www.cs.comell.edu/Information/People-/kalantar/Writings/Bahaullah/GWB/sec-24.html>).

Seguro le interesa el URL de URouLette, ¿no? He aquí cómo llegar a esa página, y hablaremos más acerca de este URL después.

[Http://www.roulette.com:8000/](http://www.roulette.com:8000/)

Pruébalo, le gustará.

Búsquedas en la Web

Si bien no existe una lista central de sitios de Web, si hay buscadores (search engines, término también traducido como máquinas de búsqueda) que le ayudan a encontrar lo que le interesa. Por ejemplo, un buscador muy popular que ya le mostré antes es Yahoo!, cuyo URL es <http://www.yahoo.com/>. Esa dirección lo lleva a una página inicial con cientos de vínculos a temas de cualquier área imaginable. Otro buscador es WebCrawler, que explora con regularidad 250 000 páginas de Web. Su URL es <http://www.webcrawler.com/>. Basta con introducir esta dirección para iniciar una búsqueda.

Por ejemplo, digamos que nos interesa encontrar información sobre estilos de aprendizaje. En la figura 3.18 se muestra cómo se introduce el término "learning styles" (porque en este caso se buscarán páginas base en inglés que contengan estas palabras, que significan "estilos de aprendizaje") en la sección de WebCrawler de la página All-In-One (Todo-En-Uno). Un clic en el botón Search (Buscar) produce los vínculos que vemos en la figura 3.19. Ahora usted puede continuar la consulta inicial visitando algunas de las 1460 páginas que WebCrawler encontró. Ahora bien, es posible que muchos de estos vínculos no sean pertinentes para su búsqueda, pero muchos sí serán. Además, los otros vínculos lo llevarán a otros lugares y le presentarán información que podría sugerir nuevas ideas y direcciones para su propio trabajo.

Yahoo! es uno de muchos buscadores que nos ayuda a encontrar información que podría estar contenida en una página base de la Web. Da la casualidad que algún ambicioso Webmeister colocó todos estos buscadores en una sola página. Usted puede encontrar la página de búsqueda All-In-One dirigiendo su navegador de la World Wide Web a <http://www.albany.net/allinone> y luego buscar por persona, tema, o incluso noticias y el tiempo.

Algunos puntos de partida interesantes

En la tabla 3.6 se muestran algunos sitios Web interesantes que tal vez usted querrá explorar para tener una idea de lo que hay allá afuera. Inclusive si usted es muy poco fanático de la Web, lea esta sección del capítulo con cuidado, o encuentre alguien que haga su tarea, trabajo, quehacer y otras obligaciones. Una vez que se conecte y comience a explorar, se le hará más que fácil pasar hora tras hora en la Web. Y la mejor forma de comenzar es a través de algunos sitios especiales como Yahoo! y de un sitio que ofrezca acceso a servidores de Web (que contienen acceso a miles de páginas base).

Sitio de la World Wide Web	URL
Yahoo en Stanford University	http://akebono.stanford.edu/yahoo
URoulette en University of Kansas	http://www.roulette.com:8000/
Una biblioteca virtual (la página base de la WWW Virtual Library)	http://info.cern.ch/hypertext/Datasources/bySubject/Overview.html
Visite el Louvre la página base del Louvre	http://mistral.enst.fr/~pioch/Louvre
Mapas y más (página base de Map Viewer)	http://pubweb.parc.xerox.com/map/
Cosas gratuitas página base del Windows Shareware Archivel	http://coyote.csum.edu/cwis/winword/winword.html
Más Casas en verdad gratuitas (página base de Free Stuff from the Internet)	http://power.globalnews.com/articles/txt/freestuf/contents.htm
Toda lo que necesita saber (página base de AskERIC Virtual Library)	http://eryx.syr.edu/COWSHOME.html
Toda la gente, todo el tiempo (página base de The Constitution of the United States (http://www.law.cornell.edu/constitution/constitution.overview.html

Tabla 3.6 Sitios de Web interesantes.

Cómo escribir la revisión de la Bibliografía

Ha llegado el momento de tomar toda la información que recopiló y organizarla de alguna manera para que comience a tener sentido. Ésta es su revisión de la bibliografía, y ahora usted tiene que sentarse y escribirla (¡qué horror!). He aquí algunas sugerencias para la redacción.

Primero, lea otras revisiones de la bibliografía. No hay argumento más convincente que el éxito. Pida a un estudiante que ya haya llevado este curso o a su profesor que le muestre una propuesta que haya sido bien recibida. Examine con detenimiento tanto el formato como el contenido de la revisión de la bibliografía. Explore también algunas de las fuentes que mencionamos en este capítulo, sobre todo las que son reseñas de la bibliografía, artículos de revistas científicas y otros trabajos de revisión.

Segundo, cree un tema unificado, o una línea de pensamiento, a seguir durante toda la revisión. No se supone que su revisión de la bibliografía sea una novela, pero casi todas las buenas revisiones se construyen partiendo de un argumento muy general hacia uno más específico y preparan el escenario para el objetivo de la investigación. Usted debe atraer al lector y crear algún interés respecto a dónde quiere dirigirse con esta investigación porque nadie ha seguido todavía ese camino.

Tercero, utilice un sistema para organizar su material. La mayor parte de las revisiones de la bibliografía se organiza cronológicamente dentro de cada tema. Por ejemplo, si usted va a estudiar las diferencias de género entre los adultos en cuanto a ansiedad y capacidad verbal, organizará todas las referencias por tema (ansiedad y capacidad verbal) y luego, dentro de cada uno de estos temas, iniciará su revisión con la referencia de fecha más antigua. De este modo, usted avanzará desde los albores hasta lo más reciente y proporcionará cierta perspectiva histórica.

Cuarto, trabaje a partir de un cuadro sinóptico. Si usted es un escritor experimentado y hábil, puede hacer caso omiso de esta sugerencia, pero si apenas está comenzando es recomendable usar esta herramienta que le ayuda a organizar la idea principal de su propuesta antes de iniciar el proceso de redacción propiamente dicho.

Quinto, construya puentes entre las diferentes áreas que revisó. Por ejemplo, si usted va a realizar un estudio intercultural para comparar la forma en que padres provenientes de diferentes culturas disciplinan a sus hijos, es posible que no encuentre mucha bibliografía sobre ese tema específico, pero ciertamente hay muchos trabajos publicados sobre la educación de los hijos en culturas individuales, y toneladas de referencias sobre la disciplina. Parte del esfuerzo creativo de escribir una propuesta es poder demostrar cómo se juntan los diferentes temas de una manera interesante y potencialmente fructífera.

Sexto, tal vez la práctica no hace al maestro, pero tampoco hace daño. Por alguna razón, la mayoría de la gente cree que una persona nace con o sin el talento para escribir. Cualquier escritor exitoso admitirá que para ser un jugador de baloncesto de primera o un consumado violinista es necesario practicar. ¿Es diferente el caso de un escritor? Si usted tiene dudas acerca de esta pregunta, encuentre un escritor profesional y pregúntele cuántas horas al día o la semana practica su oficio. Con mucha frecuencia, resultará que el escritor practica tanto como el deportista o el músico. De hecho, una escritora amiga mía da el siguiente consejo a la gente que desea escribir pero que no tiene una idea precisa del nivel de compromiso que requiere. Ella dice: "Simplemente siéntese frente a su máquina de escribir o procesador de textos y ábrase una vena". Es así de fácil.

Por tanto, la última (pero en realidad la primera) sugerencia es que practique la escritura. A medida que trabaje y averigüe dónde necesita mejorar (obtenga retroalimentación de otros estudiantes y profesores) observará sin duda un cambio positivo.

Todos los que realizan investigación comienzan en algún punto, y casi siempre una revisión bibliográfica pone las ideas y los objetivos en perspectiva. La bibliografía, y todas las herramientas que están disponibles para trabajar con la bibliografía, es su primer y mejor aliado para crear un tratamiento bien investigado y completo de lo que ha ocurrido en el pasado. Una vez identificadas las variables importantes, podrá dirigir su atención a la forma como pueden medir dichas variables, lo cual es el tema del siguiente capítulo.

CAPITULO CUATRO

Muestreo y generalizabilidad

En Neil J. Salkind "Métodos de Investigación", Prentice Hall, México, 1998.

Lo que aprenderá en este capítulo

- Por qué un buen muestreo podría ser el paso más importante de un proyecto de investigación
- La importancia de la generalizabilidad
- La diferencia entre una muestra y una población
- El significado de la palabra aleatorio
- Cómo usar una tabla de números aleatorios para seleccionar una muestra
- La diferencia entre muestreo aleatorio simple y muestreo estratificado
- La diferencia entre las estrategias de muestreo probabilísticas y no probabilísticas
- Qué es el error de muestreo y por qué es importante
- Cómo reducir el error de muestreo
- Cómo estimar el tamaño de la muestra

Imagine esto. Se le ha asignado a usted la tarea de medir la actitud general de los estudiantes de educación media hacia el registro irrestricto de sus casilleros en busca de drogas. Usted ya tiene suficiente experiencia en investigación para saber como elaborar algún tipo

de cuestionario asegurándose de que cubra las áreas de contenido importantes y sea fácil de administrar y calificar. Una vez realizado todo el trabajo preliminar, surge la pregunta más importante: ¿a quiénes le pedirá usted que contesten el cuestionario? A todos los estudiantes de todas las escuelas de educación media podrían ser millares? Imposible; sería demasiado costoso. administrar el cuestionario en aquellas escuelas en las que se ha informado de problemas con droga hacerse eso; es muy probable que también haya drogas en las escuelas que no se han identificado como escuelas problema. ¿Qué tal si se pregunta sólo a los estudiantes de último año, ya que se supone que son los que mejor conocen lo que está sucediendo en la comunidad? Tampoco puede hacerse eso porque los estudiantes de los grados inferiores también usan drogas. ¿Qué hacer?

¡Decisiones, decisiones, decisiones! Y decisiones que no se pueden tomar a la ligera éxito del proyecto depende de la forma en que se seleccione a la gente que participara en el estudio, sea que usted vaya a distribuir un cuestionario o administrar una prueba de memoria. En este capítulo examinaremos varias formas de seleccionar a los proyectos de investigación y la importancia del proceso de selección para estudio. Todo se relaciona con muestras y muestreo.

Poblaciones y muestras

En varios puntos de los primeros cuatro capítulos de este libro usted ha leído acerca de la importancia de inferir los resultados de un experimento de una muestra a una población. Ésta es la base del método inferencial. Si no es posible someter a prueba a todos los miembros de la población, la única opción es seleccionar una **muestra, un subconjunto de la población**. Una buena técnica de muestreo incluye maximizar el grado selecto representa a la población.

Una **población** es un grupo de posibles participantes al cual usted desea generalizar los resultados del estudio. Y el nombre del juego aquí es **generalizabilidad** pues para que un estudio que los resultados de una investigación tengan algún significado más allá de la situación limitada en la que se obtuvieron originalmente es necesario que se pueda generalizar desde una muestra a una población. Si los resultados son generalizables, podrán aplicarse a diferentes poblaciones que tengan las mismas características pero que estén en diferentes. Si los resultados no son generalizables (es decir, si la muestra seleccionada no es una representación exacta de la población), los resultados sólo pueden aplicarse a los integrantes de la misma muestra que participaron en la investigación original y nadie más.

Por ejemplo, supongamos que desea averiguar qué actitudes tienen los estudiantes de educación media hacia el registro de sus casilleros, en las escuelas donde se cuenta con esta ventaja para los alumnos. Se podría administrar el cuestionario a un grupo de estudiantes de último año del área de química. Pero, ¿qué tanto se parecen a la población de estudiantes que asisten a todas las escuelas de educación media del distrito no mucho. O bien, podríamos hacer las preguntas a 10% de las estudiantes de sexo femenino de primer y segundo años de bachillerato de todas las escuelas. Esta selección abarca un grupo mucho más grande que los 30 o 40 estudiantes del grupo del área de química ¿qué tan representativas son ellas? Otra vez, no mucho.

Nuestra tarea, entonces, es idear un plan que asegure que la muestra de estudiantes que seleccionemos sea representativa de *todos* los estudiantes del distrito. Si logramos este objetivo, podremos generalizar los resultados a toda la población con un alto grado de confianza, incluso si sólo usamos un porcentaje pequeño de los miles de estudiantes de bachillerato. En otras palabras, si hacemos correctamente el trabajo (seleccionar una muestra), podremos generalizar los resultados. ¿Cómo sabremos si "hicimos correctamente el trabajo"? Presentaremos algunas pautas en este capítulo, pero una forma de efectuar una especie de autoverificación es hacernos la pregunta: *¿La muestra que seleccioné de la población tiene al parecer todas las características de la población, en la misma proporción? ¿Es la muestra, efectivamente, una minipoblación?*

Para entender el proceso de muestreo, primero necesitamos distinguir entre dos tipos generales de estrategias de muestreo: probabilísticas y no probabilísticas. El **muestreo probabilístico** es un tipo de muestreo en el que se conoce la probabilidad de seleccionar un miembro individual de la población. Si hay 4500 estudiantes en todas las escuelas de bachillerato, y si 1000 de ellos están en el último año, la probabilidad de seleccionar un estudiante de último año como parte de la muestra es de $1000:4500 = 0.22$.

El muestreo **no** probabilístico es aquel en el que se desconoce la probabilidad de seleccionar cualquier miembro individual de la población. Por ejemplo, si no sabemos cuántos muchachos están inscritos en las escuelas de educación media del distrito, no podremos calcular la probabilidad de seleccionar cualquiera de ellos.

Estrategias de muestreo probabilístico

Las estrategias de muestreo probabilístico son las más utilizadas porque la selección de los participantes está determinada por el azar. Puesto que la decisión de quién entra y quién no entra en la muestra está regida por reglas no sistemáticas y aleatorias, hay una buena posibilidad de que la muestra represente verdaderamente a la población.

Muestreo aleatorio simple

El tipo más común de procedimiento de muestreo probabilístico es el muestreo aleatorio **simple**. Aquí, cada miembro de la población tiene una probabilidad *igual e independiente* de ser seleccionado como parte de la muestra. Las palabras clave aquí son *igual e independiente*. *Igual*, porque no existe alguna predisposición a escoger una persona en lugar de otra. *Independiente* porque el hecho de escoger una persona no predispone al investigador en favor o en contra de escoger otra persona dada. Si se muestrea aleatoriamente, las

características de la muestra deberán ser muy parecidas a las de la población.

Por ejemplo, ¿estaríamos realizando un muestreo aleatorio simple si escogiéramos cada quinto nombre del directorio telefónico? No, porque estaríamos violando ambos criterios de igualdad e independencia. En primer lugar, si comenzamos con el quinto nombre de la página 234 del directorio, los nombres 1, 2, 3 y 4 nunca tuvieron una probabilidad igual de ser escogidos. Segundo, si escogemos el número 5 de la lista y luego cada quinto nombre de ahí en adelante, sólo los nombres 10, 15, 20, etc., tendrán la posibilidad de ser escogidos. En otras palabras, el proceso de selección ya no es independiente.

El proceso de muestreo aleatorio simple consta de cuatro pasos:

1. Definir la población de la cual se desea seleccionar una muestra.
2. Listar todos los miembros de la población.
3. Asignar números a cada miembro de la población.
4. Aplicar un criterio para seleccionar la muestra deseada.

Por ejemplo, en la tabla 4.1 se muestra una lista de 50 nombres a los que ya se han asignado números (pasos 1, 2 y 3). No es una población muy grande pero sí excelente para fines ilustrativos. De esta población seleccionaremos una muestra de 10 individuos utilizando una **tabla de números aleatorios**. Veamos cómo funciona esto.

1. Juana	11. Susana	21. Eduardo.	31. Daniela	41. Narciso
2. Braulio	12. Nora	22. Jorge	32. Bernardo	42. Penélope
3. Enriqueta	13. Diego	23. Cecilia	33. Diana	43. Inés
4. Laura	14. Juan 5.	24. Gabriela	34. Felipe	44. Débora
5. Miguel	15. Bruno	25. María	35. Federico	45. Carla
6. Sara	16. Lorenzo	26. Celia	36. Manuel	46. Vicente
7. Teresa	17. Roberto	27. Silvestre	37. Donato	47. Gema
8. Joaquina	18. Esteban	28. Felicia	38. Eduardo M.	48. Elena
9. Jaime	19. Samuel	29. Javier	39. Timoteo	49. Alejandro
10. Tomás	20. Mario	30. Enrique	40. Manuel G.	50. Juan D.

Tabla 4.1 He aquí un grupo de 50 nombres que constituye una "población" para nuestros fines. Observe que cada uno está numerado y listo para seleccionarse.

Cómo usar una tabla de números aleatorios

Una tabla de números aleatorios es un criterio magnífico, ya que la forma como se generan los números de la tabla carece totalmente de predisposición. Por ejemplo, en la tabla 4.2 hay cantidades casi iguales de dígitos 1, 2, 3, 4, 5, etc. En consecuencia, la probabilidad de seleccionar un número que termine en 1 o en 2 o en 3, etc., es igual. Esto implica que si se relacionan nombres con los números la probabilidad de seleccionar cualquier nombre dado es también igual.

23157	48559	01837	25993
05545	50430	10537	43508
14871	03650	32404	36223
38976	49751	94051	75853
97312	17618	99755	30870
11742	69183	44339	47512
43361	82859	11016	45623
93806	04338	38268	04491
49540	31181	08429	84187
36768	76233	37948	21569

Tabla 4.2 Tabla parcial de números aleatorios. En una tabla de números aleatorios, cabe esperar que habrá cantidades iguales de los dígitos individuales, distribuidos al azar entre todos los números.

Teniendo lo anterior en mente, seleccionemos un grupo de diez nombres utilizando la tabla de números aleatorios de la tabla 4.2. Siga estos pasos.

1. Escoja un punto de partida en algún lugar de la tabla cerrando los ojos y colocando su dedo (o la punta de un lápiz) en cualquier parte de la tabla. Al seleccionar el punto de partida de este modo aseguramos que no se escogerá algún punto de partida (o nombre) específico.

Para este ejemplo, el punto de partida fue la primera columna de números, en la última fila (36768); la punta del lápiz quedó en el

cuarto dígito, el número 6.

2. El primer número de dos dígitos, entonces es 68 (en negritas en la tabla 4.3). Puesto que la población llega hasta 50, y no hay un nombre en el lugar 68, pasamos por alto este número y consideramos el siguiente número de dos dígitos. Ya que no podemos bajar más en la tabla, pasamos al tope de la siguiente columna y leemos hacia abajo, una vez más seleccionando los primeros dos dígitos. Por comodidad, hemos separado los pares de dígitos en la tabla 4.3.

23157	48559	01837	25993
05545	50430	10537	43508
14871	03650	32404	36223
38976	4975 1	94051	75853
97312	1761 8	99755	30870
11742	69183	44339	47512
43361	82859	11016	45623
93806	04338	38268	04491
49540	31 18 1	08429	84187
36768	76233	37948	21569

Tabla 4.3 Punto de partida para la selección de 10 casos utilizando la tabla de números aleatorios. Se puede comenzar en cualquier punto, en tanto ese punto se determine al azar y no se escoga intencionalmente.

3. ¡48! ¡Lo logramos! La persona 48 de la lista es Elena, y ella se convierte en la primera de la muestra de 10 miembros.

4. Si seguimos seleccionando números de dos dígitos hasta haber hallado 10 valores entre 01 y 50, habremos seleccionado los nombres de la tabla 4.1 que corresponden a los números que aparecen en negritas en la tabla 4.4. He aquí un desglose de cuáles números funcionaron y cuáles no para los fines de escoger una muestra aleatoria de 10 personas de la población de 50.

- Leyendo hacia abajo en la primera columna de números de dos dígitos, 48, 50, 03, 49 y 17 están bien, porque quedan dentro del intervalo de 01 a 50 (el tamaño de la población) y no se han seleccionado antes,
- 69 y 82 están fuera del intervalo,
- 04 y 31 están bien, y
- 76 está fuera del intervalo

Puesto que no podemos bajar más por la primera columna de números de dos dígitos, hay que subir al siguiente conjunto de números de dos dígitos (en la misma columna de cinco dígitos) en la parte superior de la columna, que comienza con el número 55.

- 55 no está dentro del intervalo,
- 43 está bien,
- 65, 75 y 61 no son aceptables,
- 18 sí,
- 85 no, y (¡por fin!)
- **33 sí, y ya tenemos las 10 personas.**

Aquí están:	Número	Nombre
	48	Elena
	50	JuanD.
	03	Enriqueta
	49	Mejandro
	17	Roberto
	04	Laura
	31	Daniela
	43	Inés
	18	Esteban
	33	Diana

Ahora ya tiene usted una muestra de 10 nombres de una población de 50, seleccionados totalmente al azar. Su muestra se escogió al azar porque la distribución de los números en la tabla parcial de números aleatorios de la tabla 4.2 se generó al azar. ¿Es una simple coincidencia que tres de los primeros cinco números (48, 50, 03, 49, 17) de la tabla parcial de números aleatorios estén agrupados? ¡Absolutamente! Este grupo de cinco es la mejor aproximación y la más representativa de cualquier muestra de cinco de toda la población, dado que cada miembro de la población tiene la misma probabilidad independiente de ser escogido.

05545	50430	10537	43508
14871	03650	32404	36223
38976	4975 1	94051	75853
97312	17618	99755	30870
11742	69183	44339	47512
43361	82859	11016	45623
93806	04338	38268	04491
49540	3118 1	08429	84187
36768	76233	37948	21569

Tabla 4.4 Selección de los números de la tabla de números aleatorios que corresponde a los 10 nombres seleccionados como muestra. Observe que los pares de números individuales están en negritas para poder distinguirlos de los que no se seleccionaron.

El gran supuesto, desde luego, es que los nombres de la población (tabla 4.1) se listaron de manera aleatoria. En otras palabras, los nombres del 01 al 20 no se listaron como los primeros 20 de los 50 porque provienen de un vecindario distinto, tienen mucho dinero, no tienen hermanos, o tienen alguna otra característica que pudiera predisponer de alguna forma la selección.

La regla general es usar un criterio que no tenga relación con lo que se está estudiando. Por ejemplo, si va a realizar un estudio sobre ofrecerse como voluntarios, ¡no va a pedir voluntarios!

Muestreo sistemático

Otro tipo de muestreo aleatorio se denomina muestreo sistemático, en el cual se escoge cada *k*-ésimo nombre de la lista. El término *k*-ésimo representa un número entre 0 y el tamaño de la muestra que usted quiere seleccionar. Por ejemplo. He aquí cómo se usaría un muestreo sistemático para escoger 10 nombres de la lista de 50 que aparece en la tabla 4.1. Para hacerlo, siga estos pasos.

1. Divida el tamaño de la población entre el tamaño de la muestra deseada. En este caso, 50 dividido entre 10 es 5. Por tanto, se escogerá cada quinto nombre de la lista.
2. Como punto de partida, escoja un nombre de la lista al azar. Haga esto usando el método de "apuntar con los ojos cerrados" o, si la lista está numerada, utilice cualquier dígito o par de dígitos del número de serie de un billete que traiga en su cartera. El billete que usamos en este ejemplo tenía como sus dos primeros dígitos 43, y éste será nuestro punto de partida.
3. Una vez determinado el punto de partida, seleccione cada quinto nombre. En este ejemplo, si usamos los nombres de la tabla 4.1 y comenzamos con Inés (#43), la muestra consistirá en Elena (#48), Enriqueta (#3), Joaquina (#8), Diego (#13), Esteban (#18), Cecilia (#23), Felicia (#28), Diana (#33) y Eduardo M. (#38).

El muestreo sistemático es más fácil y menos problemático que el aleatorio, y ésta es una razón por la que muchas veces se prefiere. El muestreo sistemático también es un poco menos deseable. Es evidente que se viola el supuesto de que cada miembro de la población tiene la misma probabilidad de ser seleccionado. Por ejemplo, dado que el punto de partida es Inés (#43), sería imposible seleccionar a Débora (#44).

Muestreo estratificado

Los dos tipos de muestreo aleatorio que acabamos de ver funcionan bien si no nos interesan características específicas de la población (como edad, sexo, grupo étnico, grupo de habilidad, etc.). En otras palabras, si se seleccionara otro conjunto de 10 nombres, supondríamos que, como ambos grupos se escogieron al azar, son efectivamente iguales. Pero, ¿qué sucede si desde un principio los individuos de la población no son "iguales"? En tal caso, querremos asegurarnos de que el perfil de la muestra coincida con el perfil de la población, y esto se hace creando una **muestra estratificada**.

La teoría en que se basa el muestreo (y todo el proceso de inferencia) es a grandes rasgos la siguiente: si usted puede seleccionar una muestra lo más representativa posible de una población, cualesquier observaciones que pueda hacer sobre esa muestra se deberán cumplir también para la población. Hasta aquí vamos bien. Sin embargo, hay ocasiones en que el muestreo aleatorio deja demasiado al azar, sobre todo si no hay garantía de que las distribuciones de los miembros de la población dentro de la muestra sean iguales. En esos casos se utiliza un muestreo estratificado.

Por ejemplo, si la población es 82% católicos, 14% protestantes y 4% judíos, la muestra deberá tener las mismas características *si la religión es una variable importante*. Es crucial entender la última parte del enunciado anterior. Si ésta o aquella característica de la población no tiene relación alguna con lo que se está estudiando, no hay razón para preocuparse por obtener una muestra que siga el mismo patrón de la población y esté estratificada según una de esas variables.

Supongamos que la lista de nombres de la tabla 4.1 representa una población *estratificada* (hombres y mujeres), y que el tema del estudio son las actitudes hacia el aborto. Puesto que las diferencias de sexo o género podrían ser importantes, usted quiere una muestra que refleje las diferencias de género de la población. La lista de 50 nombres consiste en 20 mujeres y 30 hombres, o sea, es 40% mujeres y 60%

hombres. La muestra de 10 deberá reflejar esa distribución y tener 4 mujeres y 6 hombres. He aquí cómo se seleccionaría semejante muestra utilizando muestreo aleatorio estratificado.

1. Los hombres y las mujeres se listan por separado.
2. Cada miembro de cada grupo recibe un número. En este caso, los hombres se numerarían del 01 al 30 y las mujeres del 01 al 20.
3. Usando una tabla de números aleatorios, se seleccionan 4 mujeres al azar de la lista de 20.
4. Usando una tabla de números aleatorios, se seleccionan 6 hombres al azar de la lista de 30.

Aunque podrían presentarse ejemplos sencillos (con un solo estrato o capa) como éste, en muchos casos será necesario estratificar según más de una variable. Por ejemplo, en la figura 4.1 una población de 10000 niños se estratifica según las variables de año (40% de primer año, 40% de tercer año, 20% de quinto año) y lugar de residencia (30% rural, 70% urbano). Se utiliza la misma estrategia: seleccionar 10% (puesto que 1000 es 10% de 10 000) de cada una de las capas estratificadas para producir la muestra de la figura 4.1. Por ejemplo, de los 1200 niños rurales de primer año, se seleccionó aleatoriamente 10%, o 120 niños. De forma similar, se escogieron 140 niños urbanos de quinto año.

Residencia	Año			Total
	1	3	5	
Rural	1200 (120)	1200 (120)	600 (60)	3000 (300)
Urbana	2800 (280)	2800 (280)	1400 (140)	7000 (700)
Total	4000 (400)	4000 (400)	2000 (200)	10000 (1000)

Figura 4.1 La selección de una muestra cuando hay más de un estrato consiste en tomar una proporción de cada nivel. Aquí el tamaño de la muestra se indica entre paréntesis debajo del tamaño de cada grupo de la población.

Muestreo por cúmulos

Ultimo tipo de muestreo probabilístico es el **muestreo por cúmulos, en el que se seleccionan unidades** de individuos y no los individuos mismos. Por ejemplo, usted podría estar realizando una encuesta de las actitudes de los padres hacia la vacunación. En lugar de asignar aleatoriamente padres individuales a dos grupos (digamos, aquellos a los que se les enviará material informativo y aquellos a los que no se les enviará), usted podría simplemente identificar 30 consultorios pediátricos de la ciudad y luego, utilizando una tabla de números aleatorios, seleccionar 15 para un grupo y designar 15 para el segundo grupo.

El muestreo por cúmulos ahorra mucho tiempo, pero hay que tener la seguridad de que las "unidades" (en este caso los padres que acuden a cada pediatra) sean lo suficientemente homogéneas como para que cualesquier diferencias que haya dentro de la unidad misma no contribuyan a una predisposición. Por ejemplo, si un pediatra se niega a vacunar a los niños antes de cierta edad, eso introduciría una predisposición que usted querría evitar.

Estrategias de muestreo no probabilístico

La segunda categoría principal de estrategias de muestreo, el muestreo no probabilístico; comprende aquellas en las que se desconoce la probabilidad de escoger un solo individuo. En este caso, hay que suponer que los miembros en potencia de la muestra no tienen una probabilidad igual e independiente de ser seleccionados. Examinemos algunos de estos métodos de muestreo.

Muestreo de conveniencia

El muestreo de conveniencia es justo lo que indica su nombre. Un entrenador de fútbol da a cada miembro de su equipo un cuestionario. El público (el equipo) es cautivo, y es una forma muy cómoda de generar una muestra. ¿Fácil? Sí. ¿Aleatoria? No. ¿Representativa? Tal vez, pero hasta cierto grado. Quizá usted reconozca este método de muestreo como la razón por la que tantos experimentos en psicología, realizados en Estados Unidos, se basan en resultados obtenidos usando estudiantes universitarios de segundo años; estos estudiantes son un público cautivo y a menudo deben participar para obtener los créditos de un curso.

Muestreo por cuotas

Usted podría encontrarse en una situación en la que necesitara obtener una muestra estratificada según ciertas variables, pero por alguna razón no es posible efectuar un muestreo estratificado. En este caso, el muestreo por cuotas podría ser lo mejor.

En el muestreo **por** cuotas se escogen personas con las características deseadas (digamos, niños rurales de primer año) pero no se selecciona aleatoriamente de la población un subconjunto de todos esos niños, como ocurriría en un muestreo estratificado proporcional. Más bien, el investigador continuaría reclutando niños hasta cumplir con la cuota de 120. El 1760 niño rural de primer año no tiene posibilidad de ser escogido, y ésta es la principal razón por la que esta técnica de muestreo es no probabilística.

He aquí otro ejemplo de sistema por cuotas. Digamos que usted tiene que entrevistar 20 estudiantes universitarios de primer año de ambos sexos. Primero usted podría entrevistar 10 hombres y, sabiendo que la distribución de hombres y mujeres está dividida aproximadamente 50/50, usted entrevistaría las siguientes 10 mujeres que llegaran, y con ello habría acabado. Si bien el muestreo por cuotas es mucho más fácil que el estratificado, también es menos preciso. Imagine cuánto más fácil sería encontrar 10 estudiantes de primer año de sexo masculino que 10 hombres específicos, que es lo que usted tendría que hacer si realizara un muestreo estratificado.

En la tabla 4.5 se listan los diferentes tipos de estrategias probabilísticas y no probabilísticas que hemos visto, y se indica cuándo deben usarse, junto con algunas de sus ventajas y desventajas.

Muestras y error de muestreo

Por más que se esfuerce el investigador, es imposible seleccionar una muestra que represente *perfectamente* a la población. Desde luego, el investigador podría seleccionar toda la población como muestra, pero esto elimina el propósito del muestreo: hacer una inferencia a una población con base en una muestra más pequeña.

Una forma de expresar la falta de congruencia entre la muestra y la población se expresa como el **error de muestreo**. Este error es la diferencia entre las características de la muestra y las características de la población de la cual se seleccionó la muestra. Por ejemplo, la estatura promedio de 10000 alumnos de quinto año es 102 cm. Si usted toma 25 muestras de 100 alumnos de quinto año y calcula la estatura promedio para cada grupo de 100 alumnos, tendrá un promedio para cada grupo, es decir, 25 promedios. Si todos esos promedios son exactamente 102 cm, no habrá error de muestreo. Sin embargo, es casi seguro que no se obtendrá tal resultado. La vida no es tan fácil ni la selección de muestras es así de perfecta. Lo más probable es que usted obtenga valores como 102.5, 102.2, 101.6, 100.1 cm, etc. La cantidad de variabilidad, o la dispersión, de estos valores nos da una idea de la magnitud del error de muestreo. Cuanto mayor es la diversidad de los valores de las muestras, mayor es el error.

Piense un momento en lo que sucedería si toda la población de 10 000 alumnos de quinto año filara la muestra. ¡La estatura promedio sería 102 cm! ¡Perfecto! ¡Cero error! ¿La moraleja? Cuanto mayor es la muestra, menor es el error de muestreo, porque cuanto

Tipo de muestreo	Cuándo debe usarse	Ventajas	Desventajas
Estrategias probabilísticas			
Muestreo aleatorio simple	Cuando los miembros de la población son similares	Asegura un alto grado de representatividad	Tardado y tedioso
Muestreo sistemático	Cuando los miembros de la población son todos similares	Asegura un alto grado de representatividad; no hay que usar una tabla de números aleatorios	Menos verdaderamente aleatorio que el muestreo aleatorio simple
Muestreo aleatorio estratificado	Cuando la población es de naturaleza heterogénea y contiene varios grupos distintos	Asegura un alto grado de representatividad de todos los estratos de la población	Tardado y tedioso
Muestreo por cúmulos	Cuando la población consiste en unidades más que en individuos	Fácil y cómodo	Posibilidad de que los miembros de las unidades difieran entre sí y reduzcan la efectividad del muestreo

Estrategias de muestreo no probabilísticas

Muestreo de conveniencia	Cuando la muestra es cautiva	Cómodo y económico	La representatividad es dudosa
Muestreo por cuotas	Cuando hay estratos pero no es posible un muestreo estratificado	Asegura cierto grado de representatividad de todos los estratos de la población.	La representatividad es dudosa

Tabla 4.5 Resumen de las diferentes técnicas de muestreo. Al comenzar a considerar la selección de una muestra, hay que tener en cuenta

muchos factores, todos los cuales afectarán el tipo y el tamaño de la muestra que escala.

más grande es una muestra más se aproxima su tamaño al tamaño de la población y es más representativa de esa población.

El proceso exacto para calcular el error de muestreo, que se expresa como un valor numérico, rebasa el alcance de este libro, pero el lector debe tener presente que su propósito al seleccionar una buena muestra es minimizar ese valor. Cuanto más pequeño sea el valor, menor discrepancia habrá entre la muestra y la población.

Pero hay más. Ya sabemos que cuanto más grande es una muestra, más representativa es de una población. Digamos que llega el momento de probar si hay una diferencia entre muestras. Resulta que, cuanto mejor las muestras representan su respectiva población, más exacta es la prueba de las diferencias. En otras palabras, un mejor muestreo da lugar a pruebas más exactas y verdaderas de las diferencias entre poblaciones.

¿Cómo minimizamos el error de muestreo? Sencillo. Utilizamos buenos procedimientos de selección como los que describimos al principio del capítulo y aumentamos el tamaño de la muestra hasta donde es posible y razonable. La siguiente pregunta que usted está a punto de hacer (espero) es, «¿qué tan grande debe ser la muestra?» ¡Qué bueno que preguntó! Pasemos a la última sección del capítulo para entender mejor la respuesta a esa pregunta.

¿Qué tan grande es grande?

Ahora que usted sabe algo acerca de muestreo, ¿qué tantos de esos estudiantes de educación media necesita seleccionar de la población de 4500? Si con 50 basta, ¿no es mejor 500? ¿Y por qué no 1500, si tiene el tiempo y los recursos suficientes para el proyecto?

Usted ya sabe que una muestra demasiado pequeña no es representativa de la población, y que una demasiado grande es una exageración. Muestrear demasiados estudiantes de educación media anularía el propósito del muestreo, pues ya no se estaría aprovechando la utilidad de la inferencia. Hay quienes piensan que cuanto más grande es una muestra, mejor; pero tal estrategia no es lógica desde un punto de vista económico o científico. Una muestra demasiado grande no aumenta la precisión de la prueba de la pregunta en un grado que justifique el costo y el trabajo de obtener una muestra de ese tamaño. Necesitamos un método para calcular el número *real* de estudiantes de educación media que conviene seleccionar para la muestra. Recuerde, cuanto menos representativa de la población sea la muestra, mayor será el error de muestreo presente. Además, cuanto mayor sea el error de muestreo, menos generalizables a la población serán los resultados y menos precisa será la prueba de la hipótesis nula.

Estimación del tamaño de la muestra

Es posible estimar el tamaño de muestra que se necesita para representar de la manera más exacta posible la población, pero esto va a requerir un poco de fe por parte del lector por dos razones. Primera, implica conceptos que están un poco más allá del nivel de este libro. Sin embargo, dado que el lector es inteligente y entusiasta, es probable que entienda los fundamentos de este material. Segundo, y más importante, es preciso estimar ciertos valores, lo cual muchas veces es difícil aun para los expertos.

He aquí los componentes que necesitamos para estimar los tamaños de muestra de dos grupos que se van a comparar entre sí.

s = la desviación estándar de la variable dependiente, es decir; qué tanta variabilidad hay en los puntajes para esta variable

t = el valor crítico necesario para rechazar la hipótesis nula

D = la magnitud que usted estima tiene la diferencia promedio entre los dos grupos.

Tomemos como ejemplo el caso de un estudio en el que usted está tratando de determinar si un programa de lectura es más eficaz que otro. La variable dependiente es el puntaje de lectura promedio.

La forma como estimamos la desviación estándar de una muestra antes de probar los miembros de la muestra es examinando los puntajes y estadísticas descriptivas de la variable dependiente cuando esta variable se usó en otros estudios. Por ejemplo, digamos que esta prueba de lectura tiene un manual, y que el manual indica que la desviación estándar para alumnos de primero de bachillerato en el pasado ha sido de 15.

A continuación, introducimos el valor crítico que refleja la diferencia que usted esperaría si el azar (y no cualquier tratamiento) fuera el único factor que operara. Los valores críticos se pueden encontrar en tablas especializadas. Examinaremos más a fondo este concepto en el capítulo 8.

Por último, necesitamos tener una idea de a cuánto ascenderá la diferencia entre los dos grupos. Si el puntaje que puede obtenerse en la prueba de lectura está entre 20 y 50, ciertamente no esperamos que la diferencia entre los dos promedios sea de más de 30.

Hay que ser razonables.

Para este ejemplo,

- la desviación estándar es igual a 15 (supuestamente informada en otro estudio),
- el valor t es 2 (el valor que obtuvimos de nuestra tabla especial), y
- la diferencia estimada entre los grupos es de 6, que también es una cifra de orden de magnitud.

Dados estos valores, podemos usar la siguiente fórmula para calcular n , que es el tamaño de muestra estimado necesario para cada grupo bajo las restricciones que se han establecido:

$$n = \frac{2s^2 \times t^2}{D_2}$$

donde n = tamaño de muestra estimado para cada grupo
 s = desviación estándar basada en estudios anteriores
 t = valor crítico
 D = nuestra estimación de la diferencia entre cada grupo

Si sustituimos valores, la fórmula queda así:

$$n = \frac{(2 \times 15^2) \times 2.00^2}{6^2}$$

Esta fórmula revela que, dadas las anteriores restricciones, necesitamos unas 50 personas en cada grupo.

Lo que resulta interesante examinar son los efectos de cambiar ciertos valores (como la desviación estándar y la diferencia estimada) sobre el valor estimado de n , el tamaño de muestra que necesitamos. Vea lo que sucede en la tabla 4.6 cuando cambia la diferencia estimada entre los dos grupos.

Efecto sobre el tamaño de la muestra cuando modificamos la magnitud del error estimado			
s	t	d	n
15	2	6	50
10	2	6	22
5	2	6	6

Efecto sobre el tamaño de la muestra cuando modificamos la magnitud de la diferencia estimado			
s	t	d	n
10	2	2	200
10	2	4	50
10	2	6	22

Tabla 4.6 Podemos ver lo drástico que puede ser el efecto cuando cambiamos lo diferencia estimada entre grupos.

Cuanto mayor es la diferencia estimada (de 2 a 4 a 6), menos grande necesita ser el tamaño de las muestras. ¿Por qué? Porque cuanto mayor es la diferencia entre las muestras, menos representativas tienen que ser éstas de la población para que podamos ver el efecto, si existe. Esto es un poco como decir que es más fácil percibir una diferencia grande entre dos cosas que una diferencia pequeña. En la tabla 4.6 podemos ver lo drástico que puede ser el cambio. Si duplicamos el tamaño estimado de la diferencia (de 2 a 4), reducimos drásticamente el tamaño de muestra estimado necesario, de 200 en cada grupo a 50.

¿Qué es lo más importante en cuanto al tamaño de muestra? Tenga presentes las consideraciones siguientes:

- En general, cuanto mayor es la muestra (dentro de límites razonables), menor es el error de muestreo y mejores son los resultados.
- Si va a utilizar varios subgrupos en su trabajo (como hombres y mujeres, ambos de 10 años de edad, y residentes rurales saludables y no saludables), asegúrese de que su selección inicial de sujetos sea lo bastante grande como para contemplar la división necesaria en grupos de sujetos.
- Si usted va a enviar por correo encuestas y cuestionarios (y ya sabe usted lo que puede sucederles a muchos de ellos), contemple un aumento de 40%-50% en el tamaño de muestra para dar cabida a las pérdidas en el correo y la falta de cooperación de los sujetos.
- Por último, recuerde que lo grande es bueno, pero lo apropiado es mejor. No gaste su dinero arduamente ganado ni su valioso tiempo generando muestras más grandes de lo necesario.

Aunque algunas personas podrían no estar de acuerdo con los temas que usted ha escogido para estudiar, lo que usted escoja es asunto

suyo en tanto presente una justificación razonable de lo que está haciendo. En cambio, su selección de una muestra es cuestión aparte. Existen muchas formas correctas de hacerlo, pero también está la forma incorrecta. Si usted escoge la forma incorrecta (porque es arbitrario y no sigue ningún plan), bien podría sabotear toda su labor de investigación porque sus resultados podrán carecer de generalizabilidad y, por tanto, de utilidad para la comunidad científica.

Ejercicios

1. Suponga que usted es el investigador principal de un estudio longitudinal que está rastreando las preferencias vocacionales desde el bachillerato hasta la edad adulta media. Liste los pasos que seguiría para seleccionar la muestra que usará en el estudio.
2. ¿Por qué es tan útil una tabla de números aleatorios como herramienta para asignar personas a diferentes grupos?
3. ¿Qué diferencia hay entre una estrategia de muestreo probabilístico y uno no probabilístico? Dé un ejemplo de cada una. También, cite las ventajas y desventajas de cada tipo de muestreo.
4. ¿Cuál es la forma más fácil de reducir el error de muestreo? ¿Qué relación hay entre el error de muestreo y la generalizabilidad de los resultados de un estudio? Por último, ¿qué sucede con el error de muestreo a medida que aumenta el tamaño de la muestra? ¿Por qué?
5. Si tiene una población de 10000 niños (50% hombres y 50% mujeres, 70% blancos y 30% de otras razas, y 57% de familias con un solo padre y 43% de familias con ambos padres), ¿qué pasos seguiría para seleccionar una muestra representativa de 150 niños?
6. Utilizando una tabla de números aleatorios, escoja seis nombres de la siguiente lista de 10.

Miguel
Susana
Sara
Hernán
Selina
Enriqueta
Guillermo
David
Sandra
Eduardo

¿Cuántos de los seis esperaría usted que fueran hombres, y cuántos mujeres? ¿Por qué?

7. ¿Qué implicaciones tiene usar una muestra demasiado grande o una demasiado pequeña?
8. A usted le interesa la diferencia entre niños que reciben enseñanza creativa y niños que reciben enseñanza tradicional, en cuanto a la creatividad de los niños. Se utiliza una escala de creatividad que produce un posible puntaje entre 70 y 100. Por investigaciones previas, usted estima que la desviación estándar es 15. Usted tiene acceso a 60 individuos (30 en cada grupo) y estima que la diferencia en la escala de creatividad entre los niños con enseñanza creativa y aquellos con enseñanza tradicional será de 10 puntos. Con las restricciones señaladas, ¿qué tamaño de muestra se necesita para cada grupo?
9. ¿Qué riesgos conlleva aumentar demasiado el tamaño de una muestra?
10. ¿Cuándo debe usarse muestreo por cúmulos y muestreo aleatorio simple?
11. ¿Qué es error de muestreo?

CAPITULO CINCO Medición del comportamiento

En Neil J. Salkind “Métodos de Investigación”, Prentice Hall, México, 1998.

Lo que aprenderá en este capítulo

- Por qué la medición es una parte importante del proceso de investigación
- Qué incluye el proceso de medición
- Cuáles son los diferentes niveles de medición, y cómo se aplican

- Qué es la confiabilidad
- Los diferentes tipos de confiabilidad, y cómo se usan
- Cómo aumentar la confiabilidad de una prueba
- Qué es la validez
- Los diferentes tipos de validez, y cómo se usan
- Cómo aumentar la validez de una prueba
- La relación entre confiabilidad y validez

El proceso de medición

Aun sin saberlo, es probable que usted dedique mucho tiempo a juzgar las cosas que suceden a su alrededor. En muchos casos, tales juicios son informales ("Realmente me gustó la forma en que Luis presentó ese material"), pero a veces son tan formales como es posible ("85% de sus respuestas fueron correctas").

En ambos ejemplos, se está emitiendo un juicio acerca de un resultado específico. De eso se trata el proceso de medición, y no debemos subestimar su importancia para el proceso de investigación. Toda nuestra labor y empeño para tratar de contestar ésta o aquella pregunta interesante de nada sirven si aquello que nos interesa no puede determinarse, medirse, calibrarse, valorarse, evaluarse, clasificarse, ordenarse, calificarse, graduarse, acomodarse, estimarse, puntuarse, encuestarse o ponderarse (¿capta usted la idea?).

La definición clásica de medición fue sugerida hace más de 45 años por el psicólogo experimental S.S. Stevens (1951) como la "asignación de numerales a objetos o sucesos siguiendo ciertas reglas". Sin menoscabo del respeto que nos merece el profesor Stevens, esta definición puede ampliarse para decir que la medición es la asignación de valores a resultados. Los números (como 34.89 y \$54980) son valores, pero también lo son resultados como color del pelo (rojo o negro) y clase social (baja o alta). De hecho, cualquier variable por definición puede adoptar más de un valor y puede medirse. Son estos valores los que nos interesan como parte del proceso de medición.

En este capítulo presentaremos algunos de los conceptos importantes del proceso de medición, incluidos los niveles de medición, un sistema de clasificación que ayuda a determinar qué se está midiendo y las dos cualidades primarias que debe tener cualquier herramienta de evaluación: confiabilidad y validez.

Niveles de medición

Debemos a Stevens no sólo la definición de medición en la que se basa gran parte del contenido de este capítulo, sino también un método para clasificar los diferentes resultados en lo que él llamó niveles de medición. Un nivel de medición es la escala que representa una jerarquía de precisión dentro de la cual una variable podría evaluarse. Por ejemplo, tomemos la variable estatura, que puede definirse de varias maneras, cada una de las cuales corresponde a un nivel específico de medición.

Una forma de medir la estatura consiste sencillamente en colocar a las personas en categorías como A y B, sin hacer referencia a su tamaño real en centímetros, metros o pies. Aquí, el nivel de medición es nominal porque asignamos las personas a grupos con base en la categoría a la que pertenecen.

Una segunda estrategia sería colocar a las personas en grupos rotulados a lo largo de alguna dimensión, como Alto y Bajo ("esto" o "aquello"). En este caso también estamos colocando a las personas en grupos, pero al menos hay alguna distinción más allá de un simple rótulo categórico. En otras palabras, los rótulos "alto" y "bajo" tienen algún significado en el contexto en el que se usan, mientras que categoría A y categoría B sólo nos dicen que los grupos son diferentes, sin que se conozca la naturaleza de tal diferencia. Aquí, el nivel de medición es ordinal.

Una tercera estrategia es aquella en la que vemos que Raquel es 13 centímetros más alta que Gregorio. No sólo sabemos que hay una diferencia entre ambas mediciones, sino que también conocemos la magnitud exacta de esa diferencia (13 cm). Aquí, el nivel de medición es de intervalo.

Por último, la estatura de una persona podría medirse en una escala que tiene un cero verdadero. Aunque en las ciencias sociales y de la conducta puede haber problemas con este nivel de medición de razón, tiene sus ventajas, como veremos más adelante en este capítulo.

Hay que tener presentes dos cosas en lo tocante a esta idea de nivel de medición. Primera, las cualidades de un nivel de medición (como el nominal) son características también del siguiente nivel hacia arriba. En otras palabras, las variables medidas en el nivel ordinal también contienen las cualidades de las variables medidas en el nivel nominal. Así mismo, las variables medidas en el nivel de intervalo contienen las cualidades de las variables medidas en los niveles tanto nominal como ordinal. Por ejemplo, si sabemos que Luis tiene 175 cm de estatura y Laura tiene 163 cm de estatura (nivel de medición de intervalo o posiblemente de razón), sabemos también que Luis es más alto que Laura (nivel ordinal de medición), y que Luis y Laura tienen diferente estatura (nivel nominal de medición). Segunda, en cualquier proyecto de investigación una variable de resultado pertenece a uno de estos cuatro niveles de medición. La clave, desde luego, es cómo se mide dicha variable.

A continuación presentamos un tratamiento más detallado de cada uno de estos niveles de medición, con ejemplos y aplicaciones. En la

tabla 5.1 hacemos un resumen de los cuatro niveles, y de lo que podemos y no podemos decir acerca de ellos.

<i>Nivel</i>	<i>Cualidades</i>	<i>Ejemplo</i>	<i>Qué se puede decir</i>	<i>Qué no se puede decir</i>
Nominal (<i>categorías</i>)	Asignación de rótulos	· Género (masculino o femenino) · Preferencia (gusta o no gusta) · Votación (en pro o en contra)	Cada observación pertenece a su propia categoría.	Una observación prepresenta "más" o "menos" que otra observación.
Ordinal (<i>categorías y órdenes</i>)	Asignación de valores a lo largo de alguna dimensión subyacente	· Rango universitario · Orden de llegada en una carrera	Una observación se clasifico arriba o abajo de otra.	Que tanto una variables es mayor o menor que otra.
Intervalo (<i>categorías y ordenes y tiene intervalos iguales</i>)	Distancias iguales entre puntos	· Número de palabras correctamente deletreadas · Puntajes en pruebas de inteligencia	Un puntaje difiere de otro en alguna medida que tiene intervalos equiespaciados.	La magnitud de la diferencia es una representación exacta de las diferencias en la variable que se está estudiando.
Razón (<i>categorías y ordenes y tiene intervalos iguales y un punto cero</i>)	Cero significativo y no arbitrario	· Temperatura · Edad · Peso · Tiempo	Un valor es el doble de otro o ninguna cantidad de esa variable puede existir.	No mucho

Tabla 5.1 Los niveles de medición.

Nominal

El nivel nominal de medición, de la palabra latina *nomin* (nombre), describe variables de naturaleza categórica y que difieren en calidad más que en cantidad. Es decir la variable que estamos examinando caracteriza nuestras observaciones de modo tal que cada una puede colocarse en una (y sólo una) categoría. Además, podemos rotular esas categorías a nuestro antojo. Todos los niveles nominales de medición son exclusivamente cualitativos.

Por ejemplo, el color del pelo (rubio, rojo, negro), la raza (negra o afroamericana, blanca o angloamericana, americana nativa) y filiación política (socialista, socialdemócrata, peronista, sandinista) son ejemplos de variables de nivel nominal. Incluso es posible usar números en la medición de variables de nivel nominal, aunque los números no tienen un valor intrínseco. Por ejemplo, asignar los hombres al grupo 1 y las mujeres al grupo 2, o asignar a todos los hombres de línea de un equipo de fútbol americano camisetas con los números del 40 al 50, son ejemplos de medición nominal o categórica. Los números no tienen un significado intrínseco; son sólo rótulos que identifican las cosas que se miden.

Un ejemplo de estudio que utiliza una variable de nivel nominal sería uno que examinara los méritos de dos programas escolares que buscan facilitar la integración de niños "normales" con niños que tienen discapacidades mentales severas (Cole, Vandercook y Rynders, 1988). La variable nominal o categoría en este caso es el tipo de programa en el que participaron los niños, el de Amigo Especial o el de Tutor Igual. Los niños podían participar en un programa o en otro pero no en ambos. Los investigadores examinaron las diferencias en la interacción del niño con la discapacidad y el niño sin ella en función del tipo de programa en el que participaron. Se estudiaron las diferencias en la interacción social durante el programa, durante el juego libre y durante una sesión tutorial.

Debemos recordar ciertas cosas acerca del nivel nominal de medición. Primera, las categorías son mutuamente excluyentes; no se puede estar en más de una categoría a la vez. No es posible ser al mismo tiempo judío y católico (¡aunque celebre tanto Hanukkah como Navidad!). Segunda, si se usan números como valores, no tienen significado alguno más allá de la simple clasificación. No es posible saber si alguien de la categoría 3 es menos o más inteligente que alguien de la categoría 1.

Ordinal

El nivel ordinal de medición describe variables que se pueden ordenar a lo largo de algún tipo de continuo. Las variables no sólo se pueden colocar en categorías, sino que también se pueden ordenar. Por esta razón, el nivel ordinal de medición a menudo se refiere a las variables como ordenamientos de diferentes resultados, aunque sólo se manejen dos categorías, como grande y pequeño.

Por ejemplo, ya vimos que alto y bajo son dos posibles resultados cuando se mide la estatura. Estos valores son ordinales porque reflejan un ordenamiento dentro del continuo de estatura. Asimismo, el rango de usted dentro de su generación al graduarse de bachillerato se basó (probablemente) en su promedio de calificaciones. Usted pudo ser el primero de 300, o el centésimo quincuagésimo de 300. Cabe señalar que no es posible inferir el promedio de calificaciones a partir de este ordenamiento, sino sólo la posición relativa a otros. Usted podría haberse graduado primero de 300 con un promedio de 9.54, o como el centésimo quincuagésimo de 300 con un promedio de 9.71.

A partir de alto y bajo o primero y 150-ésimo no es posible saber qué tan alto o qué tan bajo, o qué tan buen estudiante es el individuo, porque los niveles ordinales de medición no incluyen esta información. Lo que sí podemos decir es que si Diana es más baja que Juana y Juana es más baja que Lucía, entonces Diana también es más baja que Lucía. Así, pues, aunque no es posible hacer juicios absolutos (como qué tanto más alta es Lucía que Diana), sí pueden hacerse juicios relativos. Sólo podemos asignar el valor graduado con mención honorífica así como con mención honorífica con distinción y con máxima mención honorífica con distinción para distinguir aún más entre quienes se gradúan con mención honorífica. Esta escala es de naturaleza ordinal.

Intervalo

El nivel de intervalo de medición, del latín *interval lum* (que significa espacios entre paredes) describe variables que tienen intervalos iguales entre ellas (como tenían las paredes construidas por los soldados romanos). Las variables de nivel de intervalo nos permiten determinar la diferencia entre puntos a lo largo del mismo tipo de continuo que mencionamos en la descripción de la información ordinal.

Por ejemplo, la diferencia entre 30 y 40 grados es la misma que la diferencia entre 70 y 80 grados: 10 grados. Así mismo, si usted escribió correctamente 20 palabras en una prueba de ortografía y alguien más escribió correctamente 10 palabras, podemos decir con exactitud que usted escribió correctamente 10 palabras más que la otra persona. Dicho de otro modo, un grado es un grado de un grado y una palabra correctamente escrita es una palabra correctamente escrita con una palabra correctamente escrita.

Un estudio realizado por Wigfield y Eccies (1989) sobre la ansiedad ante los exámenes en unidades de educación primaria y secundaria ilustra cómo un constructo como la ansiedad se puede medir con variables de nivel de intervalo. Por ejemplo, la Escala de Ansiedad ante Exámenes para Niños (Sarason, Davidson, Lighthall, Waite y Ruebush, 1960) es una escala de 30 reactivos que evalúa diversos aspectos de la ansiedad y que produce una medida global. Reactivos como

Si faltas a la escuela y no haces una tarea, ¿qué tanto te preocupas por atrasarte respecto a los demás estudiantes cuando regresas a la escuela?

proporcionan una medida exacta del nivel de ansiedad en los niños y es una medida ampliamente utilizada de este fascinante constructo.

Para contrastar los niveles de medición de intervalo y ordinal, consideremos la variable edad donde el ordenamiento por edad es como sigue:

Mas viejo.....Mas joven
Guillermo Enriqueta Josué Raquel Julia

Sabemos que Guillermo es más viejo que Enriqueta, pero no qué tanto. En realidad, él podría ser dos años mayor que Enriqueta, y Enriqueta podría ser 20 años mayor que Josué. Las variables de nivel de intervalo nos dan esa diferencia, cosa que no pueden hacer las escalas ordinales. En términos simples, con una escala de intervalo podemos conocer la diferencia entre puntos a lo largo de un continuo (y la diferencia exacta entre las edades de Guillermo, Enriqueta, Josué, Raquel y Julia), no así con una escala ordinal.

Aunque una escala en el nivel de intervalo es más precisa y comunica más información que una escala en el nivel nominal u ordinal, debemos tener cuidado al interpretar los valores reales a lo largo de la escala. Treinta grados podrían ser 10 más que 20, y -5 podría estar a la misma distancia de +5, pero esos 10 podrían implicar una gran diferencia. Los 10 grados entre 30 y 20 podrían hacer el agua un poco más fría, pero en los 10 grados entre -5 y +5 el agua se congela. Asimismo, el hecho de que usted escribió correctamente 10 palabras más que un compañero no significa que usted tiene una ortografía dos veces mejor (2 por 10), ya que no tenemos idea de la dificultad de las palabras ni de si esas 20 palabras muestrean el universo entero de todas las palabras de la prueba de ortografía. Lo que es más importante, si usted no escribe correctamente ninguna palabra, ¿significa eso que usted no sabe escribir? Claro que no, Lo que sí significa es que en esta prueba a usted no le fue muy bien.

Razón

El nivel de razón de medición, del latín *ratio* (que significa cálculo), describe variables que tienen intervalos iguales entre ellas pero que también tienen un cero absoluto. En los términos más sencillos, esto significa que existen variables para las cuales un posible valor es cero o es posible la ausencia de la variable o rasgo.

Por ejemplo, un estudio sobre técnicas para mejorar la conducta prosocial en el salón de clases (Solomon et al., 1988) midió la conducta prosocial con puntuaciones de comportamiento. Las cinco categorías de conducta que se midieron durante un periodo de cinco años (¡qué largo!) fueron actividades cooperativas, disciplina del desarrollo, actividades que promueven el entendimiento social, actividades que destacan los valores prosociales y actividades de ayuda. Los investigadores dedicaron mucho tiempo a desarrollar sistemas que pudieran medir de manera consistente (o "confiable", que es el término que usaremos después) estos tipos de conductas. Las escalas que ellos diseñaron son por su naturaleza de razón, ya que tienen un verdadero punto cero. Por ejemplo, no es difícil imaginar que un niño no exhiba ninguna conducta prosocial.

Éste es en verdad un nivel de medición interesante, y es por mucho el más preciso. Poder decir que Samuel (quien tiene ocho años de edad) es dos veces más viejo que Erica (quien tiene cuatro años) es una forma muy precisa, aunque no la más precisa, de hablar de las diferencias entre personas en lo tocante a una variable. Imagine poder decir que la rapidez de respuesta cuando se utiliza el método A es la mitad de la que se observa cuando se usa el método B, en lugar de decir únicamente que la tasa de respuesta es más rápida (lo cual es ordinal) o 10 segundos más rápida (lo cual es intervalo).

Ésta es la escala más interesante de las cuatro, por varias razones más. Primera, el valor cero no es arbitrario. Por ejemplo, podríamos pensar que, puesto que la temperatura (en grados Celsius) tiene un punto cero (¡qué frío!) es una variable de razón. Aunque es verdad que esa escala de temperatura tiene un punto cero, se trata de un cero arbitrario. Una temperatura de cero grados Celsius no representa la ausencia de choques entre las moléculas que crean calor (la definición no técnica de temperatura, y mis disculpas a Lord Kelvin). En cambio, la escala de temperatura Kelvin sí tiene un cero absoluto teórico (cerca de -273 grados Celsius), que es el punto donde no hay actividad molecular, y es un verdadero cero o una ausencia de lo que se está midiendo (actividad molecular).

¿Por qué tanta complicación?

Seamos prácticos. En un estudio de investigación queremos medir la variable de interés de la forma más precisa posible. No sirve de mucho decir que el grupo A es más débil que el grupo B si podemos decir que los miembros del grupo A realizaron en promedio 100 "abdominales" mientras que los del grupo B sólo hicieron 75 en promedio. Tener más información incrementa la potencia y la utilidad general de las conclusiones. Imagine que usted es superintendente escolar y puede gastar \$100 000 dólares en un programa de intervención temprana. ¿No le gustaría saber cuáles programas son mejores y por cuál margen, en vez de sólo averiguar que uno es "mejor" que otro?

No obstante, a veces hay que limitarse a la cantidad de información disponible. Por ejemplo, ¿qué tal si usted quiere estudiar la relación entre la edad de los adultos y su fuerza corporal, y lo único que sabe es a qué grupo pertenece un adulto (fuerte o débil) pero no su puntaje en una prueba de fuerza? Tales limitaciones son una de las restricciones al efectuar investigaciones en el mundo real; hay que conformarse con lo que se tiene. Esas limitaciones también dan lugar a uno de los aspectos creativos de la investigación: definir las variables de tal manera que la definición maximice la utilidad de la información.

¿En qué nivel de medición encontramos la mayor parte de las variables en las ciencias sociales y del comportamiento? Probablemente en el nominal u ordinal, -mientras que la generalidad de los puntajes obtenidos en pruebas (como las de aprovechamiento) producen datos en el nivel de intervalo. No obstante, es muy dudoso si los puntajes de tales mediciones como las pruebas de inteligencia y personalidad proporcionan algo más que niveles ordinales de medición. Un niño con un IQ de 110 no es 10 puntos más inteligente que un niño con un IQ de 100, pero sí podría haber obtenido 10 puntos más. Asimismo, Cristóbal podría preferir las "chispas" de chocolate del paquete A a las del paquete B con una frecuencia dos veces mayor, pero esto no implica necesariamente que le gustan dos veces más. Ahí radica un punto importante: la forma como decidimos medir un resultado define el nivel de medición de ese resultado. Una frecuencia dos veces mayor es una variable en el nivel de razón, pero qué tanto le gustan las "chispas" de chocolate a Cristóbal es una actitud y por tanto una variable de naturaleza ordinal.

La mayoría de los investigadores se toma ciertas libertades al tratar variables ordinales (como los puntajes en una prueba de personalidad) como variables en el nivel de intervalo, y no hay problema en tanto recuerden que los intervalos podrían ser (y probablemente sean) desiguales. Al interpretar sus datos, esos investigadores deben tener en cuenta tal desigualdad.

También, hay que tener presente que la tipología de niveles de medición de Stevens no se ha escapado de ser cuestionada. En los 50 años que tiene de existencia esta metodología, han surgido varias dudas acerca de la utilidad del sistema y de qué tan bien refleja las variables del mundo real que los investigadores deben evaluar (Velleman y Wilkonson, 1993).

Primordialmente, dichas críticas se concentran en el hecho de que una variable podría no ajustarse fácilmente a ninguna de las cuatro clasificaciones, sin por ello dejar de ser valiosa. Por ejemplo, si bien la inteligencia no es una variable en el nivel de razón (nadie carece totalmente de ella), ciertamente está más allá del nivel de intervalo en sus aplicaciones de la vida real. En otras palabras, la taxonomía podría ser demasiado estricta para aplicarse a datos del mundo real. Al igual que tantas otras cosas en el mundo de la investigación, esta taxonomía de cuatro niveles es un punto de partida con el que se puede trabajar pero que no tiene que obedecerse como una ley

Confiabilidad y validez: por qué son importantes

Podemos tener el automóvil más llamativo del camino, pero si las ruedas están ovaladas podemos olvidarnos de un buen manejo y un paseo cómodo. Los neumáticos, el punto donde "el caucho toca el camino", son cruciales.

Del mismo modo, podemos tener la pregunta de investigación más imaginativa del mundo, con una hipótesis bien definida y claramente expresada, pero si las herramientas que usamos para medir el comportamiento que deseamos estudiar son defectuosas, podemos olvidarnos del éxito. La confiabilidad (o coherencia) y la validez (las cualidades de hacer lo que debe hacer) de un instrumento de medición son indispensables, ya que la ausencia de estas cualidades podría explicar por qué actuamos incorrectamente al aceptar o rechazar nuestra hipótesis de investigación.

Por ejemplo, usted podría estar estudiando el efecto de cierto programa de adiestramiento sobre las habilidades verbales de niños con un leve retraso mental, y está usando una prueba cuya confiabilidad y validez es dudosa. Supongamos por el momento que el tratamiento en verdad funciona bien y podría ser la razón de que haya diferencias significativas en las habilidades verbales de los grupos que reciben el tratamiento, en comparación con las de los grupos que no lo reciben. Puesto que el instrumento que usted está usando para evaluar las habilidades verbales no es siempre lo suficientemente sensible como para captar los cambios en la conducta verbal de los niños, puede olvidarse de detectar diferencias en sus resultados por más bueno que sea el tratamiento (y por sólida que sea su hipótesis). Con eso en mente, recuerde: las herramientas de evaluación deben ser confiables y válidas; de lo contrario, la hipótesis de investigación que usted rechace podría ser correcta ¡y usted nunca lo sabrá!

La confiabilidad y la validez son nuestra primera línea de defensa contra conclusiones espurias e incorrectas. Si el instrumento falla, todo lo demás falla también. Pasemos ahora a un tratamiento más detallado de la confiabilidad y la validez, qué son y cómo funcionan.

Una definición conceptual de la confiabilidad

Aquí vamos otra vez con otro juego de sinónimos. ¿Qué tal fiable, consistente, estable, fiel, predecible, fidedigno? ¿Capta la idea? Algo que es confiable funcionará en el futuro como lo ha hecho en el pasado. Una prueba o medida de conducta confiable puede medir la misma cosa más de una vez y producirá el mismo resultado.

Podemos usar cualquiera de los sinónimos anteriores de la palabra confiable como definición inicial, pero es importante entender primero la teoría en que se basa la confiabilidad. Por tanto, comencemos con el principio.

Cuando hablamos de confiabilidad, hablamos de puntajes. El desempeño de cualquier persona respecto a cualquier variable consiste en un puntaje formado por tres componentes claramente definidos, como se muestra en la figura 5.1.

$$\text{Puntaje observado} = \text{puntaje verdadero} + \text{Puntaje de error}$$

Error de método

Error de rasgo

Figura 5.1 Componentes de la confiabilidad. El puntaje verdadero y el puntaje de error son los componentes primarios.

Primero está el puntaje observado. Éste es el puntaje que usted registraría (u observaría) realmente en una situación de investigación. Es el número de palabras correctas en una prueba, el número de sílabas memorizadas, el tiempo que toma leer cuatro párrafos de prosa o la velocidad de respuesta. Puede ser la variable dependiente de su estudio o cualquier otra variable que se esté midiendo. Cualquier puntaje observado consiste en los otros dos componentes: puntaje verdadero y puntaje de error.

El segundo componente, el puntaje verdadero, es un reflejo perfecto del valor verdadero de esa variable, descontando cualesquier otras influencias internas o externas. En otras palabras, una persona dada tiene un solo "puntaje verdadero" respecto a una variable en particular. Si se realizan mediciones repetidas podrían obtenerse varios valores para una variable, pero el valor verdadero sólo es uno. Sin embargo, nunca podemos determinar cuál es ese valor. ¿Por qué? En primer lugar, porque la mayor parte de las variables, como memoria, inteligencia, agresión e incluso estatura (somos más altos en la mañana, ya que nuestra columna vertebral se comprime a medida que avanza el día) no se pueden medir directamente; y en segundo lugar, porque el proceso de medición es imperfecto.

No obstante, el proceso de medición siempre supone que existe un puntaje verdadero. Por ejemplo, respecto a una variable como la inteligencia, cada persona tiene un puntaje verdadero que refleja exactamente (y teóricamente) el nivel de inteligencia de esa persona. Supongamos que por alguna magia el verdadero puntaje de inteligencia de usted es 110. Si entonces se le administra a usted una prueba de inteligencia y obtiene un puntaje observado de 113, la prueba habrá sobreestimado su cociente de inteligencia. Sin embargo, dado que el puntaje verdadero es un concepto teórico, no hay forma de saber eso.

El tercer componente en la figura 5.1 es el puntaje de error, que abarca todas esas razones por las que el puntaje verdadero y el puntaje

observado difieren. Por ejemplo, Miguel podría escribir 85 de 100 palabras correctamente en una prueba de ortografía. ¿Significa esto que Miguel "tiene una ortografía 85% correcta" todos los días y en todos los exámenes de ortografía? Pues no. Lo que significa es que este día, en este examen, Miguel escribió correctamente 85 palabras de 100. Quizá mañana, con un conjunto diferente de 100 palabras, Miguel escribiría 87 o 90, o incluso 100, correctamente. Tal vez, si pudiera medirse su verdadera capacidad para escribir correctamente, sería 88. ¿Por qué las diferencias entre su puntaje verdadero (88) y su puntaje observado (85)? En una palabra, error. Tal vez no estudio tanto como debía haberlo hecho, o quizá no se sentía bien. Quizá no pudo escuchar diariamente al profesor dictar cada palabra. Tal vez las instrucciones respecto a dónde debía escribir las palabras en el formato de examen no estaban claras. Tal vez se le rompió el lápiz. Quizá, quizá, quizá.... Todas son fuentes de error.

Todas estas posibles explicaciones ponen de manifiesto que los puntajes repetidos para casi cualquier variable regularmente son diferentes entre sí, ya que el rasgo que se está evaluando cambia de momento a momento, y el instrumento que se está usando puede cambiar (aunque sea muy poco) y no es perfecto (ningún instrumento de medición lo es).

¿Qué constituye los puntos de error?

Vayamos más allá del concepto general de puntajes de error. En la figura 5.1 vemos también que los puntajes de error se componen de dos elementos que ayudan a explicar por qué difieren los puntajes verdaderos y los observados.

El primer componente de los puntajes de error se denomina error de método. Ésta es la diferencia entre el puntaje verdadero y el observado que se debe a la situación de prueba. Por ejemplo, digamos que usted está a punto de presentar un examen en su clase de introducción a la psicología. Usted ha estudiado bien, ha asistido a los repasos y siente confianza en que conoce el material. Sin embargo, cuando usted se sienta para presentar un examen, ve que hay preguntas de igualar (cuál elemento de la columna A va con qué elemento de la columna B) y preguntas tipo crucigrama, ¡y usted estaba esperando preguntas de opción múltiple! Además, las instrucciones acerca de cómo igualar elementos no están claras. En lugar de alcanzar su pleno potencial en el examen (o lograr un puntaje lo más cercano posible a su puntaje verdadero), usted obtiene un puntaje menor. El error entre los dos puntajes se debe al método de medición, las instrucciones poco claras, y demás.

El segundo componente es el error de rasgo. Aquí, la razón de la diferencia entre los puntajes verdadero y observado es característica de la persona que está presentando la prueba. Por ejemplo, si usted olvidó sus anteojos y no puede leer los problemas, o si no estudió, o si simplemente no entiende el material, la fuente de la diferencia entre el puntaje verdadero (lo que usted realmente sabe si ninguna otra cosa interviene) y lo que usted obtiene en la prueba (el puntaje observado) es resultado de errores de rasgo.

En la tabla 5.2 se listan algunos ejemplos de fuentes importantes de error que pueden afectar los puntajes en las pruebas de una prueba a otra. Cuanto más influyen estos factores, menos exacta es la medición. Es decir, cuanto más influyen los factores, menos probable es que el puntaje observado se acerque lo más posible al puntaje verdadero, que es el objetivo final.

Muy bien, usted ha sido consecuente y ha seguido leyendo esto, pero ¿Qué tienen que ver los componentes del error con la confiabilidad? Sencillamente, que cuanto más se puede acercar una prueba o instrumento de medición al puntaje verdadero, más confiable es ese instrumento. ¿Cómo nos acercamos más? Reduciendo las porciones de error de la ecuación de la figura 5.1. Entonces, conceptualmente, la confiabilidad es un cociente, como se muestra en la figura 5.2. Aquí podemos ver que, a medida que el puntaje de error se reduce, el grado de confiabilidad aumenta y se acerca a 1. En un mundo perfecto, no habría error y la confiabilidad sería 1, ya que el puntaje real sería igual al puntaje observado. Asimismo, a medida que el error aumenta, la confiabilidad disminuye porque más de lo que se observa se debe a algo que no puede predecirse con mucha exactitud: las contribuciones cambiantes de los errores de método y de rasgo.

$$\text{Confiabilidad} = \frac{\text{Puntaje verdadero}}{\text{Puntaje verdadero} + \text{Puntaje de error}}$$

Figura 5.2 Fórmula conceptual para la confiabilidad.

Fuente de error	Ejemplo
Características generales del individuo	<ul style="list-style-type: none"> · Nivel de habilidad · Habilidad para presentar pruebas · Habilidad para entender instrucciones
Características duraderas del individuo	<ul style="list-style-type: none"> · Nivel de habilidad en relación con el rasgo que se evalúa · Habilidades para presentar pruebas específicas para el tipo de reactivas de la prueba
Factores individuales temporales	<ul style="list-style-type: none"> · Salud - Fatiga

	<ul style="list-style-type: none"> · Motivación · Tensión emocional · Ambiente de la prueba
Factores que afectan la administración de la prueba	<ul style="list-style-type: none"> · Condiciones de la prueba - Interacción entre el examinador y el sujeto de la prueba · Predisposición en la calificación
Otros factores	<ul style="list-style-type: none"> · Suerte (en serio)

Tabla 5.2 Fuentes de error en la confiabilidad. El error puede ser parte del método empleada para evaluar la conducta a de la persona a rasgo que se está evaluando.

Cómo aumentar la Confiabilidad

La confiabilidad está íntimamente relacionada con el puntaje verdadero y el de error. Dado un puntaje verdadero fijo (lo que siempre es el caso, ¿no?), la confiabilidad disminuye a medida que el componente de error aumenta. Por tanto, si queremos un instrumento confiable, tenemos que reducir el error. No podemos afectar el puntaje verdadero directamente, así que minimizamos las fuentes de error externas (tener instrucciones claras y estandarizadas, traer más de un lápiz en caso de que se rompa la puntilla de uno, asegurarse de que el recinto sea cómodo) que podamos controlar. También hay que esforzarse por minimizar los errores de rasgo (hacer que los sujetos duerman bien la noche anterior, posponer la evaluación si alguien no se siente bien, etcétera).

He aquí un resumen de algunas formas importantes de aumentar la confiabilidad.

1. *Aumentar el número de reactivos u observaciones.* Cuanto mayor sea la muestra del universo de conductas que usted está investigando, más probable será que la muestra sea representativa y confiable.
2. *Elimine los reactivos poco claros.* Un reactivo poco claro no es confiable porque algunas personas responderán a él de una manera y otros responderán de forma distinta.
3. *Estandarice las condiciones en las que se administra la prueba.* Si los alumnos de cuarto año tienen que presentar una prueba de aprovechamiento mientras hay maquinaria ruidosa en operación justo afuera de la ventana del salón, o la calefacción está demasiado alta, ciertamente podemos esperar que tales condiciones afecten el desempeño, y por ende la confiabilidad.
4. *Modere la facilidad y dificultad de las pruebas.* Cualquier prueba que es demasiado difícil o demasiado fácil no refleja con exactitud el desempeño del sujeto.
5. *Minimice los efectos de sucesos externos.* Si ocurre un suceso de especial importancia, sean las vacaciones de primavera, la firma de un tratado de paz, el retiro de un miembro académico importante, etc. - cerca del momento en que se administra la prueba, posponga la evaluación. Es demasiado probable que tales sucesos acaparen la atención a expensas de un desempeño real.
- 6- *Estandarice las instrucciones.* Guillermo en una clase y Cecilia en otra deberán estar leyendo instrucciones idénticas y deberán presentar la prueba en exactamente las mismas condiciones.
- 7- *Mantenga procedimientos de calificación coherentes.* Quienquiera que haya calificado una pila de exámenes tipo ensayo le dirá que leer el primero es muy distinto de leer el último. Procure ser consistente al calificar; aunque esto implique usar una hoja que tenga puntajes en una columna y criterios en la otra.

Cómo se mide la confiabilidad

Usted conoce a los científicos: les encantan los números. No debe sorprendernos, entonces, que se utilice un concepto estadístico muy útil y fácil de entender llamado correlación (y la medida de la correlación, el coeficiente de correlación) para medir la confiabilidad. Usted aprenderá a calcular el coeficiente de correlación en el capítulo 9. Las correlaciones se expresan como un valor numérico que se representa con una r minúscula. Por ejemplo, la correlación entre la prueba 1 y la prueba 2 se representaría como

$$r_{\text{prueba1 . prueba2}}$$

Por ahora, lo único que usted necesita saber acerca de la correlación y la confiabilidad es que cuanto más similares sean los puntajes en términos de cambio de un momento a otro (de una prueba a otra), mayor será la correlación y más alta la confiabilidad. La confiabilidad es cuestión del instrumento, no del individuo.

Por ejemplo, como pronto veremos, una forma de medir la confiabilidad de una prueba es administrar la prueba a un grupo de personas en un punto dado del tiempo y luego administrar la misma prueba al mismo grupo de personas en un segundo punto del tiempo, digamos cuatro meses después.

Ahora bien, pueden suceder varias cosas cuando se tienen dos juegos de puntajes. Es posible que el puntaje de todo mundo baje del tiempo 1 al tiempo 2, o que suba. En ambos casos, si los puntajes tienden a cambiar de forma similar y en la misma dirección, la

correlación tenderá a ser positiva y la confiabilidad será alta.

Por otra parte, ¿qué tal si las personas que obtienen puntajes altos en el tiempo 1 obtienen puntajes bajos en el tiempo 2, o las personas que obtienen puntajes bajos en el tiempo 1 obtienen puntajes altos en el tiempo 2? Entonces la confiabilidad no sería buena; más bien, sería baja o inexistente porque no hay coherencia en el desempeño entre el tiempo 1 y el tiempo 2. En general, cuando los puntajes en la primera administración permanecen en la misma posición *relativa* en la segunda (altos en la prueba 1 y altos en la prueba 2, por ejemplo), la confiabilidad de la prueba será sustancial. Los coeficientes de confiabilidad (que son más o menos la misma cosa que los coeficientes de correlación) varían entre -1.00 y 1.00. Un valor de 1.00 sería una confiabilidad perfecta, donde no hay ningún error en el proceso de medición. Las pruebas estandarizadas que se usan en la mayor parte de los proyectos de investigación y que veremos en el siguiente capítulo suelen tener coeficientes de confiabilidad del orden de 0.80 a 0.90.

Tipos de confiabilidad

La confiabilidad es un concepto, pero también es una medida práctica de qué tan consistente y estable podría ser un instrumento de medición o una prueba. Hay varios tipos de confiabilidad, cada uno para un propósito distinto. A continuación estudiaremos esos tipos y la forma en que se usan. En la tabla 5.3 se muestra una comparación y un resumen de la información.

Tipo de confiabilidad	Qué es	Cómo se logra	Aspecto del valor de confiabilidad
Prueba original y repetida	Una medida de estabilidad	Administrar la misma prueba en dos puntos del tiempo distintas al mismo grupo de personas.	$r_{\text{tiempo1 tiempo2}}$
Formas paralelas	Una medida de equivalencia	Administrar dos formas diferentes al mismo grupo de personas.	$r_{\text{forma1 forma2}}$
intercalificador	Una medida de acuerdo	Hacer que dos personas califiquen las conductas y determinar el grado de acuerdo entre ellas.	Número de acuerdos/número de observaciones totales

Tabla 5.3 Diferentes tipos de confiabilidad. Sea cual sea el tipo de instrumento que usted desee usar, hay un tipo de medida de la confiabilidad apropiado para usted.

Confiabilidad de prueba original y repetida

Dos sinónimos de confiabilidad que utilizamos al principio de esta sección fueron consistencia o estabilidad. La *confiabilidad de prueba original y repetida* es una medida de qué tan estable es una prueba con el tiempo. En este caso se administra la misma prueba al mismo grupo de personas en dos puntos del tiempo distintos. En otras palabras, si se administra una prueba en el tiempo 1 y luego se vuelve a administrar en el tiempo 2 ¿serán estables los puntajes de la prueba con el tiempo? ¿El puntaje de Jaime en el tiempo 1 cambiará o será el mismo que su puntaje en el tiempo 2, *relativo* al resto del grupo?

Una pregunta importante para establecer la confiabilidad de prueba original y repetida es cuánto tiempo debe esperarse entre las pruebas. La respuesta a esta pregunta depende del uso que se le piense dar a los resultados de la prueba y también del propósito del estudio. Por ejemplo, digamos que usted está midiendo cambios en la interacción social de adultos jóvenes durante su primer año en la universidad. Usted quiere efectuar una medición de la interacción social en septiembre y luego otra en mayo, y le gustaría saber si la prueba que usa tiene confiabilidad de prueba original y repetida. Para determinar esto, usted tendría que administrar la prueba a los mismos estudiantes en el tiempo 1 (septiembre) y en el tiempo 2 (mayo) y luego correlacionar el conjunto de puntajes. Puesto que a usted no le interesa el cambio en la interacción social a lo largo de un periodo de dos semanas, establecer la confiabilidad de prueba original y repetida en un periodo tan corto no sería útil, dada la intención del estudio.

Confiabilidad de formas paralelas

Otra forma común de confiabilidad es la de **formas paralelas o equivalencia**. En este caso, se administran diferentes formas de la misma prueba al mismo grupo de personas. Luego se correlacionan los dos conjuntos de puntajes. Se dice que las pruebas son equivalentes si la correlación es estadísticamente significativa.

¿Cuándo querría usted utilizar la confiabilidad de formas paralelas, suponiendo que ha creado (o tiene) dos formas de la misma prueba? El

ejemplo más común es cuando se necesita administrar dos pruebas del mismo constructo dentro de un tiempo relativamente corto y se desea eliminar la influencia de los efectos de la práctica sobre los puntajes de los participantes.

Por ejemplo, digamos que usted está estudiando la memoria a corto plazo. Usted lee una lista de palabras a los sujetos y les pide recitar lo que puedan recordar diez minutos después. Tal vez sería necesario repetir este tipo de prueba cada día durante siete días, pero ciertamente no podría usarse la misma lista de palabras cada día. Si así se hiciera, para el último día los sujetos sin duda habrían memorizado una buena parte de la lista como resultado de la repetición, y la prueba proporcionaría poca información acerca de la memoria a corto plazo. Lo que usted haría es diseñar varios conjuntos de palabras que en su opinión son equivalentes entre sí; luego, si pudiera establecer que son formas paralelas de la misma prueba, podrá usarlas cualquier día y esperar que los resultados del Día 1 sean equivalentes a los resultados del Día 2.

Confiabilidad intercalificadores

La confiabilidad de prueba original y repetida y la de formas paralelas son medidas de qué tan coherente es una prueba a lo largo del tiempo (prueba original y repetida) y de una forma a otra (formas paralelas). Otra forma de confiabilidad es la confiabilidad intercalificadores.

La **confiabilidad intercalificadores** es una medida de la consistencia de un calificador a otro, más que de un tiempo a otro o de una prueba a otra. Por ejemplo, digamos que usted está realizando un estudio que mide la agresión en niños de edad preescolar. Como parte del estudio, usted está adiestrando a varios de sus colegas para recabar datos con exactitud. Usted ha ideado una "escala de calificación" que consiste en una lista de diferentes conductas en las que participan los niños, numeradas del 1 al 5, cada una de las cuales representa un tipo distinto de comportamiento, como se muestra en la tabla 5.4.

Conducta	Código	Definición
Hablar	1	Interacción verbal con otro niño
Juego solitario	2	Jugar solo sin interactuar con otro niño
Juego paralelo	3	Jugar unto con otros niños en lo mismo actividad
Golpear 1	4	Golpear físicamente a otro niño sin provocación
Golpear 2	5	Golpear físicamente a otro niño con provocación

Tabla 5.4 Las conductas pueden agruparse en categorías y luego utilizarse para registrar objetivamente su frecuencia, pero la confiabilidad es tan importante para este caso como para cualquier otra clase de medición.

Como puede ver, la conducta número 1 de la lista se denomina *hablar* y se define como «interacción verbal con otro niño». La conducta 4 de la lista se denomina *golpear 1* y se define como "golpear físicamente otro niño sin provocación". No están muy complicadas las definiciones, ¿verdad? Parecen funcionales y objetivas. ¿Pero quién puede saber si incluso con estas definiciones Esteban y Andrea (los dos calificadores) van a clasificar de manera idéntica los comportamientos que observan? ¿Qué tal si Andrea ve que Julia golpea a Isabel y clasifica esa conducta como número 4, pero Esteban la clasifica como número 5 porque vio a Isabel golpear primero a Julia? Estamos en problemas. Los calificadores deben ser capaces de calificar y colocar los sucesos en la misma categoría.

Para tener la seguridad de que todos los calificadores están de acuerdo, es preciso establecer la confiabilidad intercalificadores. Ello se logra haciendo que los calificadores basen el puntaje en un comportamiento y examinando después el porcentaje de acuerdo entre ellos. Digamos que usted hace que Andrea y Esteban califiquen la conducta de un niño cada 10 segundos mientras usted los adiestra en el uso de la escala de calificación. Su patrón de elecciones podría ser parecido al que se muestra en la tabla 5.5. Para calcular su confiabilidad intercalificadores, basta con tomar el número de acuerdos y dividirlo entre el total de periodos calificados (20 en este ejemplo). En su calificación antes del adiestramiento se obtiene una confiabilidad intercalificadores de 15 (el número de acuerdos) dividido entre 20 (el número de posibles acuerdos), que es .75 o 75%. Después del adiestramiento, como puede verse, el valor ha aumentado a 18/20 = .90 **90%**, que es muy respetable.

		Antes del adiestramiento																			
Periodo		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Andrea		5	5	4	5	5	3	4	2	3	2	2	3	3	3	5	3	3	4	3	4
Esteban		4	5	4	4	5	3	5	2	3	2	2	3	3	3	4	3	3	5	3	4
		Después del adiestramiento																			
Periodo		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Andrea		5	5	4	5	5	3	5	2	3	2	2	3	3	3	5	3	3	5		3 4
Esteban		4	5	4	4	5	3	5	2	3	2	2	3	3	3	5	3	3	5		3 4

Tabla 5.5 Confiabilidad intercalificadores antes y después del adiestramiento. El adiestramiento es la clave si se pone atención a las

aspectos importantes de la observación.

¿En qué consistió el adiestramiento? El director del proyecto probablemente detectó dónde estaban los errores de clasificación equivocada, modificó la definición de las conductas y analizó ejemplos con los calificadores. En la tabla 5.5 puede verse que el problema más frecuente eran desacuerdos entre las calificaciones de una conducta número 4 y una número 5, que son tipos de conductas de golpear. Aquí es donde se aclararían las diferencias entre los juicios de los calificadores.

Las consecuencias de una confiabilidad intercalificadores baja pueden ser graves. Si uno de los calificadores clasifica mal 20% de las ocurrencias, esto significa que 20% de sus datos podría ser erróneo, y eso podría echar todo a perder.

Consistencia interna

Aunque la **consistencia interna** es una forma de confiabilidad que se establece con menos frecuencia, usted, como investigador principiante, debe saber de qué se trata. La consistencia interna examina qué tan unificados están los reactivos en una prueba o evaluación. Por ejemplo, si usted está administrando una prueba de personalidad que contiene 100 reactivos distintos, es deseable que todos esos reactivos estén relacionados entre sí en la medida en que el modelo o teoría en que se basa la prueba considere que cada uno de los 100 reactivos refleja el mismo constructo de personalidad básico. Del mismo modo, si usted va a administrar una prueba de 100 reactivos dividida en cinco diferentes subescalas, cada una de las cuales consiste en 20 reactivos, cabría esperar que la prueba tenga consistencia interna si los 20 reactivos dentro de cada subescala tienen una relación más cercana unos con otros que con los reactivos de cualquiera de las otras cuatro subescalas. Si así es, cada una de las subescalas tiene consistencia interna.

La consistencia interna se evalúa correlacionando el desempeño en cada uno de los reactivos de una prueba o escala con el desempeño total en la prueba o escala, y adopta la forma de un coeficiente de correlación.

Cómo establecer la confiabilidad: un ejemplo

Uno de los mejores lugares para buscar estudios de confiabilidad es en el anuario *The Mental Measurements Yearbook* (Conoley & Kramei; 1989), un compendio de resúmenes y reseñas de pruebas que están disponibles. En esas reseñas, a menudo se describe y se comenta la forma en que se estableció la confiabilidad.

Por ejemplo, la *Multidimensional Aptitude Battery* es una prueba general de aptitud o inteligencia para adultos que se puede calificar objetivamente y que adopta la forma de diez puntajes de subpruebas, cinco verbales y cinco de desempeño. Los autores de la prueba calcularon varios tipos de confiabilidad, incluidos coeficientes de correlación de prueba original y repetida, que variaron entre .83 y .97 para la escala verbal de la prueba y entre .87 y .94 para la escala de desempeño. Los autores también calcularon otros índices de confiabilidad que dan cierta indicación de qué tan homogéneas o unidimensionales son las distintas pruebas. En otras palabras, se determinó que las pruebas evaluaban de manera consistente sólo una dimensión de aptitud o inteligencia. Si bien los resultados de estos estudios de confiabilidad no nos emocionan terriblemente (aunque sí a los autores de la prueba), es información crucial que un usuario en potencia tiene necesidad de conocer y que el autor de cualquier prueba necesita establecer para que la prueba sea útil.

Validez

¿Recuerda que al principio del capítulo mencionamos dos características indispensables de una buena prueba? La primera es que sea confiable, y es lo que acabamos de ver; la segunda es que sea válida, o que la prueba mida lo que se supone que debe medir.

Definición conceptual de validez

¿Recuerda la consistencia, la estabilidad y la predecibilidad (entre otros sinónimos de confiabilidad)? ¿Qué tal *veracidad*, *exactitud*, *autenticidad*, *genuinidad* y *solidez* como sinónimos de validez? Estos términos describen de qué se trata la **validez**: de que la prueba o el instrumento que se está usando realmente mida lo que usted necesita medir.

Cuando usted ve el término *validez*, una o más de tres cosas le deberán venir a la mente acerca de la definición y el uso del término. "Tenga presente que la validez de un instrumento a menudo se define dentro del contexto de cómo se está usando la prueba. He aquí los tres aspectos de la validez.

Primero, la validez se refiere a los resultados de una prueba y no a la prueba misma. Por tanto, si tenemos la prueba ABC de habilidades sociales, los resultados de la prueba podrían ser válidos para medir la interacción social en adolescentes. Hablamos de validez no a la luz de los resultados de una prueba.

Segundo, al igual que la confiabilidad (aunque la validez no se cuantifica tan fácilmente), la validez nunca es una cuestión de "todo o nada". Los resultados de una prueba no son simplemente "válidos" o "no válidos". Esta progresión ocurre en grados desde escasa validez hasta mucha validez.

Tercero, la validez de los resultados de una prueba se debe interpretar dentro del contexto en el que ocurre la prueba. Si no filera así,

cualquier cosa podría considerarse válida con sólo darle otro nombre. Por ejemplo, he aquí un reactivo de una prueba de 100 reactivos.

$$2+2=?$$

Casi todos nosotros reconoceríamos que esta pregunta tiene validez como medida de la habilidad para sumar. Pero si utilizamos la pregunta en un experimento que se concentra en las habilidades para multiplicar, el reactivo pierde su validez de inmediato.

La forma de examinar la validez de una prueba, entonces, es determinar si la prueba se concentra en los resultados de un estudio y si los resultados se entienden dentro del contexto del propósito de la investigación.

Al igual que con la confiabilidad, hay varios tipos de validez que usted encontrará en sus actividades de investigación. Desde luego, usted tendrá que considerar la validez cuando llegue el momento de seleccionar los instrumentos que piense usar para medir la variable dependiente que le interesa.

En la tabla 5.6 se presenta un resumen de los diferentes tipos de validez, qué significan y cómo se establecen.

Tipo de validez	¿Qué es?	¿Cómo se Logra?
De contenido	Una medida de qué tan bien las reactivas representan el universo entero de reactivas.	Pregunte a un experta si los reactivas evalúan la que usted quiere que evalúen.
De criterio Concurrente	Una medida de qué tan bien una prueba estima un criterio.	Seleccione un criterio y correlacione los puntajes en la prueba con los puntales en el criterio en el presente.
Predictiva	Una medida de qué tan bien una prueba predice un criterio.	Seleccione un criterio y correlacione los puntajes en la prueba con los puntales en el criterio en el futuro.
De constructo	Una medida de qué tan bien una prueba evalúa algún constructo subyacente.	Evalúe el constructo subyacente en el que se basa la prueba y correlacione estos puntajes con los de la prueba.
De consistencia interna	Una medida de la consistencia con que cada reactivo mide el mismo constructo subyacente.	Correlacione el desempeño en cada reactivo con el puntaje total de la prueba.

Tabla 5.6 Tipos de validez. Al igual que en el caso de la confiabilidad, hay un tipo apropiado para su investigación.

Tipos de validez

Hay cuatro tipos de validez, cada uno de los cuales sirve para establecer la veracidad de los resultados de una prueba o una herramienta de evaluación.

Validez de contenido

El tipo de validez más directo y sencillo es la **validez de contenido**. La validez de contenido es el grado en que una prueba representa el universo de reactivos del cual se extrajo y es útil sobre todo para evaluar la utilidad de las pruebas de aprovechamiento o pruebas que muestrean un área de conocimientos en particular.

¿Por qué sólo una muestra? Porque es imposible crear todos los reactivos que podrían escribirse. ¡Nada más piense en la magnitud de la tarea! Imagine escribir todos los posibles reactivos de opción múltiple sobre el material cubierto (no necesariamente contenido) en un libro de introducción a la psicología. Debe haber un millón de reactivos que concebiblemente podrían escribirse sobre los dominios de personalidad, percepción, o nada más la personalidad. ¡Podría cansarse nada más de pensar en ello! Es por eso que muestreamos de entre todos los reactivos que podrían escribirse.

Pero regresemos al mundo real. Digamos que usted se está ocupando de los cursos de historia de segundo de secundaria y la unidad trata el descubrimiento de Norteamérica y los viajes y peripecias de varios grandes exploradores europeos. Si usted fuera a crear un examen de historia con preguntas acerca de este periodo y quisiera establecer la validez de las preguntas, podría mostrárselas a un experto en la

historia de América y preguntarle:

"¿Estas preguntas representan con justicia el universo o dominio de la historia de América en la Era de los descubrimientos?" No es necesario que usted utilice palabras domingueras como "universo" y "dominio", pero sí necesita saber si ha cubierto lo que necesita cubrir.

Si sus preguntas son apropiadas, ya tiene la muestra de preguntas que prueba los conocimientos de un estudiante de segundo de secundaria en el área de la historia de América en la Era de los descubrimientos. Felicitaciones. Eso es validez de contenido.

Validez de criterio

La **validez de criterio** se ocupa de qué tan bien una prueba estima el desempeño actual (llamada **validez concurrente**) así como de qué tan bien predice el desempeño futuro (llamada **validez predictiva**). La validez de criterio es una medida del grado en que una prueba está relacionada con algún criterio. Es de suponer que el criterio con el que se está comparando la prueba tiene algún valor intrínseco como medida de algún rasgo o característica. La validez de criterio generalmente sirve para evaluar la validez de las pruebas de capacidad (habilidades actuales) y de aptitud (habilidades potenciales).

En ambos tipos de validez de criterio, se usa un criterio como medida de confirmación. Por ejemplo, digamos que usted desea investigar el uso de las calificaciones en los estudios de posgrado para predecir cuáles integrantes del programa de psicología clínica van a tener mucho éxito como investigadores. Para ello, usted localiza una muestra de "buenos" investigadores (tomando como criterio para definir "bueno" el número de artículos publicados en revistas científicas en los últimos 20 años). Luego, usted averiguaría qué calificaciones obtuvieron esos investigadores cuando eran estudiantes de posgrado, y qué tan bien su desempeño académico (o calificaciones) predijo su pertenencia al grupo de investigadores "buenos". Quizá sería conveniente también localizar un grupo de investigadores "malos" (aquellos que no han publicado nada) y determinar qué tan bien las calificaciones en sus estudios de posgrado predijeron su pertenencia al grupo de los "malos". En este caso, las calificaciones en los estudios de posgrado tendrían validez predictiva (del éxito como investigador) si dichas calificaciones (la prueba) tienen una buena correlación con el desempeño como investigador (el criterio).

Esto suena bonito y muy claro, pero ¿quién va a juzgar el valor del criterio? ¿El número de artículos publicados hace que un investigador sea eficaz? ¿Qué tal si 90% de los artículos publicados por un investigador aparecen en una revista que tiene una tasa de rechazo de 50%, y otro investigador ha publicado un solo artículo pero en una revista cuya tasa de rechazo es de 90%? ¿Y qué tal si ese único artículo que alguien publica tiene un efecto significativo y profundo sobre la dirección que seguirán las investigaciones futuras en esa disciplina? Al igual que con cualquier otro bloque de construcción del proceso de investigación, el criterio que se usa para establecer la validez se debe escoger con alguna justificación. En este caso, habría que proporcionar una justificación para suponer que el número de artículos publicados, sin importar su calidad, es lo importante (si eso es lo que usted cree).

Otro problema que se presenta con la validez tanto concurrente como predictiva es la duda grave sobre qué miden realmente las pruebas. Suponemos que si las pruebas tienen correlación con el criterio, la relación debe ser significativa. Entonces, si los resultados de la prueba de inteligencia que usted administra se correlacionan con el color de los ojos o el tamaño de la nariz o las irregularidades en la forma del cráneo, ¿significa eso que usted tiene una prueba con validez de criterio? La respuesta es afirmativa, si cree que el color de los ojos y el tamaño de la nariz y el estudio de la forma del cráneo (llamado frenología, por cierto) son buenos indicadores de la inteligencia. No se ría; la historia de la ciencia está plagada de tales supuestos y conclusiones bien intencionados (y algunos no tan bien intencionados) pero equivocados.

Validez de constructo

La validez de constructo es la más importante. Es un tipo de validez que requiere mucho tiempo y a veces esfuerzo para establecerse, pero también es la más deseable. ¿Por qué? Primero demos una definición: la **validez de constructo** es el grado en que los resultados de una prueba se relacionan con constructos psicológicos subyacentes. Esta validez vincula los componentes prácticos del puntaje de una prueba con alguna teoría o modelo de conducta subyacente.

Por ejemplo, la validez de constructo nos permite decir que una prueba que se dice es una "prueba de inteligencia" realmente mide la inteligencia. ¿Cómo se establece esta validez? Digamos que, con base en una teoría de la inteligencia (que se ha sometido a cierto escrutinio y pruebas y que ha resistido la prueba del tiempo), la inteligencia consiste en conductas tales como memoria, comprensión, pensamiento lógico, habilidades espaciales y razonamiento. Es decir, la inteligencia es un constructo representado por un grupo de variables relacionadas entre sí. Si usted desarrolla un conjunto de reactivos de prueba con base en ese constructo, y puede demostrar que los reactivos reflejan el contenido del constructo, habrá comenzado a establecer la validez de constructo de la prueba.

El primer paso para crear una prueba que tiene validez de constructo, entonces, es establecer la validez (en los términos científicos más generales) del constructo subyacente en el que la prueba se basará. Este paso podría requerir estudios y más estudios, y años de investigación. Una vez demostrada la validez del constructo, puede iniciarse el diseño de una prueba que refleje el constructo.

Hay varias formas de establecer la validez de constructo.

Primero, al igual que con la validez de criterio, podemos buscar la correlación entre la prueba que estamos creando y alguna prueba

establecida que ya se ha demostrado que posee validez de constructo. Este problema es un poco como el de "la gallina y el huevo", ya que siempre existe la duda de cómo se estableció la validez de constructo de la primera prueba.

Segundo, podemos demostrar que los puntajes de la prueba que se está diseñando diferirán entre grupos de personas que poseen y carecen de ciertos rasgos o características. Por ejemplo, si usted está desarrollando una prueba de la agresión, tal vez desearía los resultados para personas que se sabe son agresivas con los de personas que se sabe no lo son.

Tercero, podemos analizar los requisitos de tarea de los reactivos y ver si son congruentes con la teoría en que se basó la creación de la prueba. Si su teoría de la inteligencia dice que la memoria es importante, es de esperar que algunos reactivos de la prueba ejerciten esa capacidad.

Establecimiento de la validez: un ejemplo

Hablando de inteligencia, he aquí cómo tres investigadores (Krohn, Lampí y Phelps, 1988) hicieron para explorar la validez de constructo de la relativamente nueva Batería de Evaluación Kaufman para Niños (K-ABC, Kaufman Assessment Battery for Children).

El problema que estos tres investigadores atacaron es muy conocido: ¿una prueba que es válida para un grupo de personas (niños de raza blanca en edad preescolar) es válida para otro grupo (niños de raza negra en edad preescolar)? Para contestar esta pregunta, los investigadores utilizaron la que tal vez sea la estrategia más común para establecer la validez de constructo. Ellos examinaron la correlación entre la prueba en cuestión y alguna otra medida establecida y válida de la inteligencia. En este caso, se usó la Escala de Inteligencia Stanford-Binet, que es la prueba de inteligencia más ampliamente utilizada con niños pequeños.

Espero que usted se esté preguntando: "Si existe una prueba de inteligencia ampliamente utilizada y supuestamente buena, ¿qué caso tiene crear otra?" Muy buena pregunta. La respuesta es que los creadores de la prueba K-ABC (Kaufman y Kaufman, 1983) creen que la inteligencia debe explorar las capacidades cognitivas más que lo que han hecho las pruebas anteriores. La K-ABC mide tanto la inteligencia como el aprovechamiento y se basa en una orientación teórica ligada menos con la cultura que pruebas como la Stanford-Binet y la Escala Wechsler de Inteligencia para Niños (conocida como WISC).

En el estudio de Krohn, Lampí y Phelps (1988), los investigadores administraron a los mismos niños tanto la K-ABC como la Stanford-Binet y encontraron que la K-ABC era apropiada como medida de inteligencia en la población de niños negros en edad preescolar de la cual se escogió la muestra.

La relación entre confiabilidad y validez

La relación entre confiabilidad y validez es directa y fácil de entender y se expresa más o menos así: una prueba puede ser confiable, pero no válida, pero una prueba no puede ser válida si no es confiable. En otras palabras, la confiabilidad es una condición necesaria, pero no suficiente, para la validez.

Por ejemplo, regresemos a esa prueba de 100 reactivos. He aquí el mismo ejemplo que usamos antes:

$$2+2=?$$

Ahora bien, podemos garantizar que éste es un reactivo confiable porque es probable que produzca una evaluación consistente de los conocimientos de suma elemental de la persona que presenta la prueba. Pero, ¿qué tal si decimos que la prueba es de ortografía? Es obvio que el reactivo no prueba la ortografía y ciertamente no es válido como tal. Esta falta de validez, empero, no afecta su confiabilidad.

Éste podría parecer un ejemplo extremo, pero se cumple en toda el área de evaluación de la conducta. Una prueba puede ser confiable y evaluar de manera consistente algún resultado, pero a menos que ese resultado se relacione directamente con el aspecto que se está estudiando, la prueba no será válida. ¡Y no hay más que decir!

Para concluir

El proceso de medición es increíblemente importante y, al igual que tantas otras cosas que guían las labores de los investigadores, no es sencillo. Se trata de un área de trabajo plena de controversias e ideas nuevas. Permítaseme sembrar una idea en la mente del lector que ilustra lo generativo y lleno de potencial que es el estudio de la medición del comportamiento humano.

En un artículo reciente en la prestigiosa revista científica *Science*, Michelle Lampí y sus colegas (Lampí *et al.*, 1992), realizaron un estudio que fue sugerido implícitamente por el comentario de un amigo sobre la rapidez con que el bebé de Lampí estaba creciendo (como el informe de su mamá -sí, de la madre de usted, el lector- a su abuela: "¡se estiró de un día para otro!"). Los doctores suelen registrar la estatura y el peso de los bebés cada mes al principio y luego cada varios meses a medida que crecen. Pues bien, estos investigadores decidieron ver si los bebés realmente crecen en rachas especialmente rápidas, así que midieron el crecimiento de varios bebés durante un periodo largo. ¿Qué averiguaron? Seguramente se sorprenderá al enterarse de que algunos niños crecen hasta una pulgada completa (2.5

cm) ¡en un periodo de 24 horas! ¿Por qué tanto alboroto? Bueno, la longitud promedio de los bebés a esa edad es de unas 20 pulgadas (50 cm), y el cambio representa un incremento de cerca de 5%. Si usted es un hombre adulto de estatura, digamos, de 175 cm, y creciera 5% en un día, mediría 184 cm al día siguiente, y si usted es una mujer de estatura de, digamos, 162 cm, llegaría a 170 cm. Por cierto, acerca de esos pantalones que se quería comprar...

La moraleja es que sin duda hay miles de cosas que están sucediendo en las ciencias sociales y del comportamiento de las que no nos percatamos ya sea porque no las medimos correctamente (no a propósito, sino porque así es como se ha medido X o Y antes) o porque estamos haciendo supuestos incorrectos (como que el crecimiento de los bebés es un proceso continuo y uniforme, sin cambios abruptos). Lo más importante es que lo que los investigadores conocen acerca del comportamiento humano depende en última instancia de cómo miden lo que les interesa estudiar. En otras palabras, la técnica de medición empleada y las preguntas que se hacen van de la mano y están íntimamente relacionadas, tanto en lo sustancial como en lo metodológico.

¿Quiere ahorrarse pasos en su investigación? No lo haga, pero si tiene que hacerlo, no sacrifique ningún aspecto del proceso de medición.

No hay discusión; el proceso de medición es un componente crítico de cualquier proyecto de investigación e indispensable para llevarlo a buen término. Esta parte del proceso de investigación tiene especial importancia, ya que una prueba sin los niveles apropiados de confiabilidad y validez de nada le sirve a usted ni a nadie más. El empleo de herramientas de medición mal diseñadas hace que no podamos saber si vamos por buen camino o si nunca medimos con exactitud lo que queremos medir. Utilice el sentido común y busque instrumentos que ya se haya demostrado que tienen niveles respetables de confiabilidad y validez. Esto le ahorrará tiempo, esfuerzo y un sinnúmero de dolores de cabeza.

Ejercicios

1. Identifique el nivel de medición asociado a cada una de las variables que se listan aquí:

- puntaje de prueba de ortografía
- vecindario
- edad en años
- color expresado como longitud de onda
- promedio decimal de calificaciones
- color de objetos de estímulo
- tiempo al correr los 100 metros
- selección de club para después de la escuela
- puntaje de IQ
- año escolar

2. Indique cuáles de las fuentes de error en la confiabilidad son de rasgo (r) y cuáles son de método (m).

- no dormir lo suficiente la noche anterior a la prueba
- instrucciones deficientes en la prueba
- vigilante que se pasea demasiado durante la prueba
- instrucciones mal impresas y difíciles de leer
- edad

3. Describa dos formas de establecer la confiabilidad de una prueba, y explique el propósito de cada una.

4. Usted acaba de crear la prueba de historia ABC, que contiene 100 reactivos y prueba los conocimientos de historia de un estudiante. ¿Qué clase de validez necesitaría establecer para una prueba como ésta, y cómo la establecería?

5. Mencione algunas pruebas que usted haya presentado y que se haya supuesto que tenían validez predictiva.

6. Defina la frase "nivel de medición."

7. Mencione los cuatro niveles de medición, y dé un ejemplo de cada uno.

8. ¿Qué relación hay entre confiabilidad y validez?

¿Quiere saber más?

Si las obras que aquí se recomiendan no están disponibles en las bibliotecas de las instituciones de su comunidad, intente poner en práctica sus habilidades de búsqueda por Internet (véase el capítulo 3).

Métodos para medir el comportamiento

En Neil J. Salkind "Métodos de Investigación", Prentice Hall, México, 1998.

Lo que aprenderá en este capítulo

- El uso de diferentes métodos para medir el comportamiento y recabar datos
- Qué es una prueba
- Cómo se diseñan diferentes tipos de pruebas para evaluar diferentes tipos de conductas
- El empleo de pruebas de aprovechamiento en las ciencias sociales y del comportamiento
- El diseño de reactivos de opción múltiple
- Cómo efectuar un análisis de reactivos
- La aplicación de escalas de actitud
- La diferencia entre una escala de actitud de Thurstone y una de Likert
- Cómo se diseñan los cuestionarios, y cómo funcionan
- Cómo utilizar entrevistas para obtener información detallada

En el capítulo 5 usted recibió una buena dosis de los aspectos teóricos en los que se basa la ciencia de la medición, y aprendió por qué la medición es crucial para el proceso de investigación, cómo se definen la confiabilidad y la validez, y cómo pueden establecerse ambas.

En este capítulo usted comenzará a aprender a aplicar algunos de esos principios al leer acerca de los diferentes métodos que pueden servir para medir el comportamiento, que incluyen las pruebas, los cuestionarios, las entrevistas y otras técnicas.

Mientras lee este capítulo, tenga en mente varias cosas. Su preocupación primordial es decidir qué método usará para medir la conducta de interés debe ser si la herramienta que piensa utilizar es confiable y válida. Esto se aplica tanto a la prueba mejor diseñada como a ¡la entrevista aparentemente más informal. Si su prueba no "funciona", casi ninguna otra cosa funcionará.

Segundo, la forma en que usted hace su pregunta determina la manera como usted mide la(s) variable(s) de interés. Si usted quiere averiguar qué opinión tiene la gente acerca de una cuestión específica, está hablando de escalas de actitud. Si desea saber cuánta información tienen las personas acerca de un tema específico, está hablando de una prueba de aprovechamiento o alguna otra medida del conocimiento. El foco de un estudio (como los efectos de las guarderías) podría ser el mismo, ya sea que usted mida la actitud o el aprovechamiento, pero lo que usted use para evaluar su variable de resultado depende de la pregunta que hace. Es necesario decidir qué intención tiene la actividad de investigación, lo que a su vez refleja la pregunta e hipótesis de investigación originales.

Por último, tenga presente que los métodos varían ampliamente en cuanto al tiempo que toma aprender a usarlos, en cuanto al proceso de medición mismo, y en cuanto a lo que se puede hacer con la información una vez que se ha obtenido. Por ejemplo, una entrevista podría ser apropiada para determinar qué opinión tienen los maestros de los cambios en la administración de la escuela, pero no si lo que nos interesa evaluar es la fortaleza física.

Así pues, vamos a presentar un panorama general de varias herramientas de medición. A' igual que cualquier otra herramienta, use bien la que escoja, y las recompensas serán abundantes. Por lo mismo, si usa la herramienta de manera incorrecta, es posible que no pueda realizar la tarea e, incluso si logra, la calidad y el valor de su informe final serán menores que lo esperado.

¿Qué mejor lugar para comenzar que con el método de medición al que todos hemos estado expuestos una y otra vez: la bendita prueba.

Las pruebas y su creación

En los términos más generales, el propósito de una prueba es medir la naturaleza y el grado de diferencias individuales. Por ejemplo, tal vez usted deseará evaluar qué tanto conocen los adolescentes sobre cómo se contrae el SIDA. O bien podría interesarle las diferencias que hay en alguna medida de la personalidad, como el Indicador de Tipo Myers-Briggs o una prueba de inteligencia como la Escala de Inteligencia Adulta Wechsler. Las pruebas también son instrumentos que distinguen entre las personas según medidas como su tiempo de reacción o su fuerza y agilidad físicas o la estrategia que alguien escoge para resolver un problema. No todas las pruebas utilizan lápiz y papel, y la técnica que un investigador emplea para evaluar una conducta a menudo refleja su creatividad.

Una buena prueba debe poder diferenciar una persona de otra de manera confiable con base en sus puntajes. Antes de continuar, cabe hacer una aclaración. Usaremos la palabra prueba en todo este capítulo para indicar una herramienta o técnica que sirve para evaluar el comportamiento, no como sinónimo de variable dependiente. Si bien podemos usar una prueba para evaluar algún resultado, también podríamos usarla para formar categorías. Supongamos que usted quiere investigar la efectividad de la terapia del comportamiento y la medicación sobre los desórdenes compulsivos. Usted podría usar una prueba para clasificar los sujetos como severa o levemente enfermos y luego aplicar otra prueba para evaluar la efectividad de cada tratamiento.

¿Por qué usar pruebas?

Las pruebas son muy populares para la evaluación de resultados sociales y comportamientos porque tienen un propósito muy específico. Las pruebas producen un puntaje que refleja el desempeño respecto a alguna variable (como inteligencia, tiempo de reacción o nivel de actividad) y pueden satisfacer diversas necesidades del investigador; como se resume en la tabla 6.1.

Que hacen	Cómo lo hacen	Un ejemplo
Ayudan a los investigadores a determinar el resultado de un estudio	Las pruebas se usan como variables dependientes	un investigador desea saber cuál de dos programas de capacitación es más eficaz.
Ayudan a proporcionar información de diagnóstico y preselección	Las pruebas por lo regular se administran al principio de un programa para tener una idea del status de la persona.	Un maestro necesita saber en qué tipo de programa de lectura debe colocar a un niño dado.
Ayudan en el proceso de ubicación	Las pruebas sirven para ubicar a la gente en distintos entornos con base en ciertas características.	Un trabajador social en el área de la salud mental necesita ubicar un cliente en un programa de rehabilitación por uso de drogas.
Ayudan en la selección	Las pruebas sirven para decidir qué personas se admitirán o ciertos programas.	Un comité de estudios de pos grado uso los puntales de pruebas para tomar decisiones sobre la admisión de estudiantes de licenciatura
Ayudan a evaluar resultados	Los pruebas sirven para determinar si se alcanzaron las metas de un programa.	Un superintendente escolar utiliza una encuesta para medir si los programas de servicio interno han tenido algún impacto sobre las actitudes de los maestros.

Tabla 6.1 Qué hacen las pruebas, y cómo lo hacen.

En primer lugar, y lo más importante, las pruebas ayudan a los investigadores a determinar el resultado de un experimento. Sencillamente, las pruebas son la vara con la que se mide la efectividad de un tratamiento o se evalúa el estado de una variable como estatura o preferencia electoral en una muestra. Puesto que los resultados de las pruebas nos ayudan a determinar el valor de un experimento, también pueden ayudarnos a construir y probar hipótesis.

Segundo, las pruebas pueden utilizarse como herramientas de diagnóstico y preselección, pues nos dan una idea de las cualidades y defectos de un individuo. Por ejemplo, la Prueba de Preselección por Desarrollo Denver (DDST, Denver Developmental Screening) evalúa el desarrollo lingüístico, social, físico y personal de niños pequeños. Aunque en el mejor de los casos ésta es una prueba de preselección general, proporciona información importante acerca del status de un niño y de las áreas de su desarrollo que podrían requerir atención.

Tercero, las pruebas ayudan a colocar. Por ejemplo, los niños que por pocos días no tienen la edad requerida para inscribirse en el jardín de niños podrían presentar una batería de pruebas para determinar si tienen las habilidades necesarias para entrar ya en la escuela. En Estados Unidos, por ejemplo, los estudiantes de educación media a menudo se inscriben en cursos de colocación avanzada y luego presentan exámenes para quedar exentos de cursos universitarios básicos requeridos. En ambos casos, los puntajes de las pruebas son útiles para hacer una recomendación respecto a la etapa de un programa en la que alguien debe colocarse.

Cuarto, las pruebas ayudan a seleccionar. Es común que la admisión a estudios de pos-grado la determine, al menos en parte, el puntaje obtenido por el solicitante en pruebas especiales. Las empresas con frecuencia realizan exámenes para seleccionar los individuos que van a contratar, a fin de asegurarse de que cuentan con las habilidades básicas necesarias para terminar la capacitación y desempeñarse de forma competente.

Por último, se usan pruebas para evaluar los resultados de un programa. No es sino hasta que recabamos información relacionada con la pregunta que hicimos y actuamos con base en esa información, que realmente sabemos si el programa que estamos evaluando tuvo, por ejemplo, el impacto esperado. Si a usted le interesa evaluar la efectividad de un programa psicoterapéutico sobre la depresión, es poco probable que pueda juzgarla sin algún tipo de evaluación formal.

Qué aspecto tienen las pruebas

Tal vez las pruebas más conocidas para usted sean las de aprovechamiento, que por lo regular incluyen reactivos de opción múltiple como

La raíz cúbica de 8 es

- a) 2
- b) 4
- c) 6
- d) 8

Las preguntas de opción múltiple son reactivos comunes en muchas de las pruebas que usted presentará en toda su carrera universitaria. Sin embargo, las pruebas pueden asumir muy diversos aspectos, sobre todo cuando es preciso satisfacer las necesidades de las personas que están siendo probadas y muestrear el comportamiento que interesa conocer mejor.

Por ejemplo, no es razonable esperar que personas con un impedimento visual grave presenten una prueba con lápiz y papel en la que tienen que rellenar pequeños círculos situados unos muy cerca de otros. Asimismo, si queremos averiguar más acerca de las interacciones sociales de los niños con otros de su misma edad, es probable que lo mejor sea observarlos jugar y no preguntarles acerca de sus juegos.

Después de hacer tales consideraciones, es necesario decidir qué forma podría asumir una prueba. Algunas de las preguntas que surgen al decidir cómo debe ser una prueba y cómo debe administrarse son las siguientes:

- ¿Se administrará la prueba empleando lápiz y papel o de otra manera?
- ¿Cuál es la naturaleza de la conducta que se evaluará (cognoscitiva, social, física)?
- ¿Las personas informan de su propia conducta (llenando un formato de informe sobre el mismo que contesta) o se observa su conducta?
- ¿La prueba tiene un límite de tiempo?
- ¿Las respuestas a los reactivos son subjetivas (y su calificación es un tanto arbitraria) u objetivas (y hay reglas bien definidas para calificarlas)?
- ¿La prueba se administra en grupo o individualmente?
- ¿Se pide a quienes presentan la prueba reconocer la respuesta correcta (como en una prueba de opción múltiple) o proporcionarían (como en un reactivo de llenar los huecos o una pregunta abierta)?

Tipos de pruebas

Las pruebas se diseñan para un propósito específico: evaluar un resultado cuyo valor distingue diferentes individuos entre sí. Puesto que hay muchos tipos distintos de resultados que podrían medirse, hay diferentes tipos de pruebas que podrían servir. Por ejemplo, si usted quiere saber qué tan bien un grupo de estudiantes de último año de bachillerato entendió una lección de física reciente, lo apropiado sería una prueba de aprovechamiento.

Por otra parte, si lo que le interesa es entender mejor la estructura de la personalidad de un individuo, sería más apropiada una prueba como el Inventario de Personalidad Multifásico Minnesota o la Prueba de Apercepción Temática, dos pruebas de personalidad muy populares pero también muy distintas.

Lo que sigue es un análisis de algunos de los principales tipos de pruebas con los que usted se topará en sus trabajos de investigación, que distinguen unos de otros y cuál es la mejor manera de utilizarlos.

Pruebas de aprovechamiento

Las pruebas de aprovechamiento sirven para medir los conocimientos en un área específica, y son las más comúnmente utilizadas cuando el resultado que se está midiendo es el aprendizaje. Estas pruebas también sirven para medir la efectividad de la enseñanza que acompañó al aprendizaje. Por ejemplo, los distritos escolares a veces utilizan los puntajes obtenidos por los estudiantes en pruebas de aprovechamiento para evaluar la efectividad de los maestros.

Las pruebas de ortografía que usted presentó en la escuela primaria, su examen final de español en secundaria o su examen bimestral de química en bachillerato eran todas pruebas de aprovechamiento por la misma razón: estaban diseñadas para evaluar qué tan bien usted entendió una información específica. Hay pruebas de aprovechamiento de todos sabores, desde la prueba común de opción múltiple hasta los exámenes de falso-verdadero y de escribir un ensayo. Todas tienen sus puntos fuertes y débiles.

Existen básicamente dos tipos de pruebas de aprovechamiento, las estandarizadas y las creadas por el investigador. Las pruebas estandarizadas suelen ser producidas por editoras comerciales y tienen una amplia aplicación en muchas situaciones distintas. Lo que distingue una prueba estandarizada de otras es que viene con un conjunto de instrucciones y procedimientos de calificación que son estándar. Por ejemplo, la Prueba de Competencia Mínima de Kansas es una prueba estandarizada que se ha administrado a más de un millón de niños del estado de Kansas en ambientes rurales y urbanos, de muy diferentes clases sociales, tamaños de escuela y antecedentes. Otro ejemplo es la Prueba de Aprovechamiento de California (CAT, California Achievement Test), una prueba

estandarizada de aprovechamiento en las áreas de lectura, lenguaje y aritmética que se utiliza en todo Estados Unidos.

Las pruebas hechas por el investigador, en cambio, se diseñan para un propósito mucho más específico y están limitadas en su aplicación a un número mucho más pequeño de personas. Por ejemplo, la prueba que usted tal vez presentará en este curso seguramente estará hecha por el investigador (o profesor) y estará diseñada específicamente para el contenido de este curso. Otro ejemplo sería una prueba, diseñada por un investigador, para determinar si el uso de máquinas de enseñanza en lugar de la enseñanza tradicional afecta el aprendizaje de un idioma extranjero.

Las pruebas de aprovechamiento también pueden dividirse en otras dos categorías. Tanto las pruebas estandarizadas como las creadas por el investigador pueden estar referidas a una norma o referidas a un criterio.

Una prueba referida a una norma es una en la que se puede comparar el desempeño de un individuo en la prueba con el desempeño de otros individuos en esa prueba. Por ejemplo, si un estudiante de ocho años de edad recibe una calificación de 56 en una prueba de matemáticas, es posible usar las normas que se proporcionan junto con la prueba para determinar la ubicación de ese niño relativa a otros estudiantes de la misma edad. Las pruebas estandarizadas suelen venir acompañadas por normas, pero esto no es siempre el caso, ni la existencia de normas es una condición necesaria para considerar estandarizada una prueba. Recuerde, una prueba sólo es estandarizada si tiene un conjunto estándar o común de procedimientos de administración y calificación.

Una prueba referida a un criterio (término acuñado por el psicólogo Robert Glaser en 1963) es una en la que se define un criterio o nivel de desempeño específico, y lo único que importa es el desempeño del individuo sin hacer comparaciones con el desempeño de otros. En este caso, el desempeño se define como una función del conocimiento de algún dominio de contenido. Por ejemplo, si usted especifica un conjunto de objetivos para el curso de historia de último año de bachillerato y establece que los estudiantes, para aprobar, deben mostrar que dominan 90% de esos objetivos, estará implicando que el criterio es un conocimiento de 90%. Dado que este tipo de prueba se enfoca realmente hacia el conocimiento del contenido en un nivel específico, se conoce también como prueba referida al contenido.

¿Cuándo usar cada tipo de prueba? En primer lugar, hay que tomar esta decisión antes de comenzar a diseñar una prueba o buscar la prueba que se usará en la investigación. La pregunta básica que usted debe contestar es si le interesa conocer qué tan buen desempeño tiene un individuo en relación con el de los demás (en cuyo caso necesitará normas para efectuar la comparación) o qué tan bien el individuo domina un área de contenido dada (en cuyo caso el dominio se reflejará en el criterio empleado).

Segundo, cualquier prueba de aprovechamiento, sea cual sea su contenido, queda en una de las cuatro casillas que se muestran en la tabla 6.2, que ilustra las dos dimensiones que acabamos de describir: ¿La prueba compara los resultados con otros individuos o con algún criterio? y ¿quién diseñó o escribió la prueba?

¿La prueba compara los resultados con otros individuos o con algún criterio?

¿Quién diseñó la prueba?	Referida a una norma	Referrida a un criterio
Estandarizada	Este tipo de prueba trae un conjunto estándar de instrucciones para su administración y calificación, y el desempeño se compara con el de otro grupo similar de individuos.	Este tipo de prueba trae un conjunto estándar de instrucciones para su administración y calificación, y está diseñada de manera tal que el desempeño individual se puede comparar con cierto criterio.
Hecha por el investigador	Este tipo de prueba sirve para evaluar el desempeño dentro de algún dominio del conocimiento relativamente estrecho, comparándolo con el de otros individuos.	Este tipo de prueba sirve para evaluar el desempeño dentro de algún dominio del conocimiento relativamente estrecho, comparándolo con algún criterio predefinido.

Tabla 6.2 Clasificación de las pruebas de aprovechamiento como referidas a normas o a criterios y coma estandarizadas o diseñadas por el investigador

Reactivos de aprovechamiento de opción múltiple

¿Recuerda esas interminables horas rellenas de óvalos en hojas para lector óptico o encerrando en un círculo las A, B, C y D, adivinando cuál respuesta podría ser correcta o no, mientras se le recomendaba no adivinar si no tenía idea de cuál era la respuesta correcta?

Todas esas experiencias son parte de las preguntas de opción múltiple, por mucho el tipo de pregunta más ampliamente utilizado en

pruebas de aprovechamiento, y uno que merece especial atención.

¿Cómo es un reactivo de Opción múltiple?

Una pregunta de opción múltiple tiene su propia anatomía especial, como se muestra en la figura 6.1. Primero tenemos el enunciado cuyo propósito es preparar la pregunta o plantear pregunta de opción el problema. Luego viene el conjunto de opciones. Una de estas opciones debe ser la respuesta correcta (la alternativa A en este ejemplo), mientras que las otras tres (en este ejemplo) deben actuar como distractores.

12. Las pruebas de inteligencia administradas a niños de edad preescolar

- A. favorecen a las niñas de clase media.
- B. tienen dudosa validez de constructo.
- C. son deficientes en cuanto a habilidades motoras.
- D. no son en absoluto divertidas.

Figura 6.1 Ejemplo de reactiva de opción múltiple.

Un buen distractor es lo suficientemente atractivo como para que a una persona que no conoce la respuesta correcta le parezca plausible. La persona que presenta la prueba descarta fácilmente los distractores demasiado alejados de la realidad (como la alternativa D en la figura 6.1), y esto contribuye a la falta de validez y confiabilidad de la prueba. ¿Por qué? Porque la presencia de distractores deficientes dificulta todavía más obtener con la prueba una estimación exacta del puntaje real de la persona que presenta la prueba.

¿Usarlas o no usarlas?

Las preguntas de opción múltiple son ideales para evaluar el nivel de conocimiento que un individuo tiene acerca de un dominio de contenido específico como economía nacional, desarrollo infantil, geología, química, latín, fibras ópticas, costura o volibol. Sea cual sea el contenido de la prueba, los reactivos deben escribirse teniendo en mente los objetivos originales de las lecciones, capítulos, artículos, conferencias y otras enseñanzas a las que se haya expuesto a quienes presentan la prueba. Si su profesor de Geología 1 no tenía como objetivo la distinción entre los diferentes tipos de formas del terreno, la prueba no debe incluir reactivos sobre dicha distinción. En otras palabras, el contenido de una prueba de opción múltiple debe reflejar inequívocamente el contenido y los objetivos del material del cual se extrajeron los reactivos.

El empleo de reactivos de opción múltiple en una prueba de aprovechamiento tiene varias ventajas y desventajas, las cuales deben tenerse en cuenta si se piensa utilizar una prueba de este tipo para evaluar un resultado basado en conocimientos. He aquí algunas de las ventajas:

- Pueden servir para evaluar casi cualquier dominio de contenido.
- Son relativamente fáciles de calificar y se pueden calificar con facilidad con una máquina.
- Quienes presentan la prueba no tienen que escribir respuestas elaboradas sino sólo escoger una de las opciones de cada reactivo
- Puesto que los reactivos de opción múltiple se concentran en los conocimientos y no en la escritura, la gente que no es buena para escribir no es necesariamente castigada por no poder manifestar lo que sabe.
- Los buenos reactivos pueden utilizarse otras veces en el futuro, lo que ahorra tiempo de preparación.
- Si los distractores son buenos le pueden ayudar al profesor a diagnosticar la razón por la que quien presentó la prueba no dio la respuesta correcta.
- Es difícil fingir que se sabe cuando no se sabe, ya que las probabilidades (.25 en el caso de cuatro alternativas que incluyen una respuesta correcta) están en contra de ello.

Por otra parte, no hay que olvidar que el empleo de reactivos de opción múltiple también tiene desventajas:

- Limitan las opciones de los estudiantes para generar respuestas creativas.
- No hay oportunidad de practicar redacción.
- A algunas personas simplemente no les gustan y suelen obtener malos resultados con ellas.
- Una pregunta de opción múltiple limita el tipo de contenido que puede evaluarse.
- Los reactivos deben prepararse con cuidado, pues de lo contrario los estudiantes inteligentes detectarán las opciones mal preparadas y las eliminarán como distractores viables.

Análisis de reactivos: cómo saber si sus reactivos funcionan

Un reactivo de opción múltiple es bueno si hace una cosa muy bien: discriminar entre los que conocen la información de la prueba y los que no la conocen. Por ejemplo, un reactivo que todo mundo contesta correctamente no sirve de nada porque no le dice al examinador quién conoce el material y quién no. Por lo mismo, un reactivo que todo mundo contesta mal no proporciona información sobre el entendimiento del material por parte de quien presenta la prueba. En otras palabras, y en ambos casos, el reactivo no discrimina.

¿No sería bonito que hubiera algún índice numérico que nos dijera qué tan bueno es realmente un reactivo de opción múltiple? ¡Su deseo se ha cumplido! el análisis de reactivos genera dos índices de ese tipo: el nivel de dificultad y el nivel de discriminación. Con estas potentes herramientas, podemos evaluar fácilmente el valor de un reactivo y determinar si conviene mantenerlo en la reserva de reactivos (la colección de reactivos de opción múltiple en un área de contenido específica), modificarlo o tirarlo a la basura.

Las dos medidas que los psicométricos (qué nombre tan elegante para quienes estudian la medición, ¿no?) usan son el índice de dificultad y el índice de discriminación, dos medidas independientes pero complementarias de la efectividad de un reactivo individual. Antes de calcular cualquiera de estos índices, es preciso dividir el total de los puntajes obtenidos en la prueba en un grupo "alto" y un grupo "bajo". Para crear estos dos grupos, siga estos pasos:

1. Coloque todos los puntajes obtenidos en orden descendente, de modo que el puntaje más alto esté a la cabeza de la lista.
2. Considere como grupo alto 27% superior de los puntajes.
3. Considere como grupo bajo 27% inferior de los puntajes.

Por ejemplo, si usted tiene 150 adultos en su muestra, los 41 puntajes más altos (27% de 150) estarían en el grupo alto, y los 41 puntajes más bajos estarían en el grupo bajo.

¿Por qué 27%? Ese número es mágico porque es la cantidad que maximiza la discriminación entre los dos grupos. Si se acuerda, lo que queremos es calcular los índices de dificultad y discriminación para contrastar los grupos de personas que obtienen buenos resultados y que obtienen malos resultados.

Para cada reactivo, examine el número de alternativas que se escogieron construyendo una tabla como la que se muestra en la tabla 6.3. Por ejemplo, 23 personas del grupo alto seleccionaron la alternativa A (que es la respuesta correcta) y 6 personas del grupo bajo seleccionaron la alternativa D.

- Las pruebas de inteligencia administradas a niños de edad preescolar

- A. favorecen a los niños de clase media.
- B. tienen dudosa validez de constructo.
- C. son deficientes en cuanto a habilidades motoras.
- D. están muy lejos de ser divertidas.

Alternativa	A	B	C	D	E
Grupo alto(n=41)	23	12	4	2	41
Grupobajo(n=41)	11	9	15	6	41
Total	34	21	19	8	82

Índice de dificultad = .41
 Índice de discriminación = .29

Tabla 6.3 Datos para calcular las índices de dificultad y de discriminación.

El **índice de dificultad** no es más que la proporción de examinados que contestaron correctamente el reactivo. La fórmula es

$$D = \frac{NCh + NCj}{T}$$

donde D = nivel de dificultad

NCh = número de personas del grupo alto que contestó correctamente el reactivo
 NCj = número de personas del grupo bajo que contestó correctamente el reactivo
 T = Total de personas en los grupos alto y bajo

En este ejemplo, el nivel de dificultad es

$$D = \frac{23 + 11}{82} = .41$$

lo que significa que el nivel de dificultad promedio para ese reactivo es .41, 041%, un reactivo moderadamente difícil. (Si todos contestaran mal el reactivo, el nivel de dificultad sería de 0%.)

El índice de discriminación es un poco más complicado: es la fracción de los examinados del grupo alto que contestaron correctamente el reactivo menos la fracción de los examinados del grupo bajo que contestó correctamente el reactivo. Este valor puede variar entre -1.00 y +1.00. Un índice de discriminación de 1.00 indica que el reactivo discrimina a la perfección; todos los integrantes del grupo alto

contestaron en forma correcta y todos los integrantes del grupo bajo contestaron de modo incorrecto. Por lo mismo, si el índice es -1.00, esto significa que todos los del grupo bajo lo contestaron correctamente, y nadie del grupo alto contestó de manera correcta (que no es precisamente lo que debía suceder).

Para calcular el índice de discriminación, utilice esta fórmula:

$$d = \frac{NCh - NC_1}{(.5)T}$$

donde: d = nivel de discriminación

NCh = número de personas del grupo alto que contestó correctamente el reactivo

NC₁ = número de personas del grupo bajo que contestó correctamente el reactivo

T = Total de personas en los grupos alto y bajo

En este ejemplo, el nivel de discriminación es

$$d = \frac{23 - 11}{(.5)82} = .29$$

o sea, 29%. Lo que queremos son reactivos que discriminen entre los que "saben" y los que "no saben", pero que no sean ni demasiado fáciles ni demasiado difíciles. En la figura 6.2 se muestra cómo los reactivos sólo pueden discriminar perfectamente cuando el nivel de dificultad es de 50%. A medida que la dificultad aumenta o disminuye, la discriminación se ve restringida. Podemos modificar el nivel tanto de discriminación como de dificultad en un esfuerzo por mejorar los reactivos.

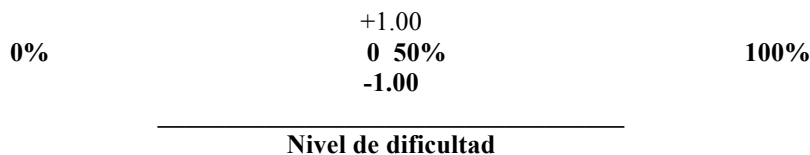


Figura 6.2 Relación entre las índices de dificultad y de discriminación de un reactiva.

Para modificar el nivel de dificultad, trate de aumentar o disminuir el atractivo de las alternativas. Si modifica el atractivo de las alternativas, verá que también cambia el índice discriminación. Por ejemplo, si una alternativa incorrecta se hace más atractiva, es probable que discrimine de manera más efectiva porque engañará a quienes casi saben la respuesta correcta.

Aunque calcular estos índices puede ser una tarea laboriosa, son prácticamente la única forma de averiguar si un reactivo está funcionando como debe. Muchas personas que utilizan con regularidad reactivos de opción múltiple sugieren hacer lo siguiente para llevar el control de sus reactivos.

Cada vez que escriba un reactivo, hágalo en una tarjeta común de unos 8 cm por 13 cm. En la parte de atrás de la tarjeta anote la fecha de la administración de la prueba (y cualquier otra información que considere importante). Debajo de la fecha agregue cualesquier comentarios que se le ocurran y también los índices de dificultad y discriminación para ese reactivo en particular. Luego, cuando trabaje con estos reactivos en el futuro, irá creando un archivo de reactivos con diferentes grados de dificultad y niveles de discriminación. Estos reactivos se podrán reutilizar o alterar según las necesidades.

Para poder discriminar al máximo entre grupos, trate de ajustar el nivel de dificultad del reactivo (que en gran medida está bajo el control del investigador) de modo que se acerque lo más posible a la marca de 50%.

Pruebas de actitud

En tanto que las pruebas de aprovechamiento son quizá el tipo de prueba más utilizado en nuestra sociedad, en diversas aplicaciones de investigación se utilizan otros tipos.

Uno de esos tipos es la prueba de actitud, que evalúa las opiniones de un individuo acerca de un objeto, persona o suceso. Se usan escalas de actitud cuando lo que interesa es conocer qué opina alguien acerca de una cosa en particular, sea la marca de rosetas de maíz para horno de microondas que prefiere o lo que piensa acerca de las leyes sobre la eutanasia.

Por ejemplo, la figura 6.3 ilustra el formato básico de una escala de actitud sencilla. Se presenta una afirmación y luego el individuo indica su actitud según alguna escala como "De acuerdo", "Indiferente" y "En desacuerdo". La selección de reactivos a incluir y el diseño de la escala tienen sus bemoles y no deben emprenderse a la ligera. Examinemos dos de las metodologías estándar que se usaron para

crear dos tipos de escala, las escalas de Thurstone y de Likert, y veamos cómo se desarrollaron.

Reactivo	De acuerdo	Sin opinión definida	En desacuerdo
La víspera del Día de Acción de Gracias debería ser feriado.			
Los exámenes finales deberían ser opcionales.			
El comedor debería servir comida gourmet.			
Mis padres no aprecian mi inteligencia.			
Mis profesores no aprecian mi inteligencia.			

Figura 6.3 Escala simple de actitud.

Escalas de Thurstone

L.L. Thurstone fue un psicólogo famoso que creó la escala de Thurstone, un método para medir actitudes. Él razonó que si era posible averiguar qué valor asignaban los expertos a un conjunto de afirmaciones, éstas podrían ajustarse a una escala. Las respuestas de las personas a dichas afirmaciones indicarían su actitud hacia el reactivo en cuestión. Éstos son los pasos a seguir para crear una escala semejante:

1. Se escribe el mayor número posible de afirmaciones como reactivos en potencia. Por ejemplo, si lo que interesa son las actitudes de los padres hacia la escuela, algunos reactivos podrían ser
 - a. Me gusta la forma en que el maestro de mi hijo lo saluda en la mañana.
 - b. El director no se comunica de manera efectiva con los maestros.
 - c. La educación y el potencial de mi hijo están en riesgo.
 - d. Los almuerzos escolares son saludables y nutritivos.
2. Jueces, que saben mucho acerca del área de interés, colocan las afirmaciones en 11 pilas (físicamente distintas) que van desde la afirmación menos favorable hasta la más favorable. La pila 6 (que está justo a la mitad) representa una afirmación neutral. Por ejemplo, el reactivo C del punto 1 se podría calificar como 1, 2, 3, 4, o 5, porque parece ser un tanto desfavorable.
3. Las afirmaciones calificadas de manera más consistente (con baja variabilidad) por los jueces reciben la calificación promedio según su colocación. Por ejemplo, si el reactivo A se calificara como 90/10 (muy favorable), recibiría un valor de escala de 9.5.
4. Se selecciona un grupo de afirmaciones que cubra toda la gama, desde desfavorable hasta favorable. Ésa es la escala de actitud.

Una de las principales ventajas de las escalas tipo Thurstone es que se acercan lo más posible al nivel de medición de intervalo, ya que los jueces que calificaron los reactivos los colocaron en pilas que están (es de suponer) a distancias iguales entre puntos que reflejan diferencias psicológicas. Es por esta razón que la escala de Thurstone también se conocen como intervalos aparentemente iguales.

Se pide a los encuestados que marquen las afirmaciones con las que están de acuerdo. Puesto que se conoce el valor de escala que se asignó a los reactivos marcados, es fácil calcular un puntaje de actitud. Si una persona marca muchos reactivos cuyos valores de escala no son similares, o bien la actitud de ese individuo no es consistente o no está bien formada, o la escala no se creó correctamente.

Por ejemplo, he aquí algunos reactivos referentes a actitudes hacia la iglesia, en Estados Unidos, tomados de la obra clásica de Thurstone sobre medición de actitudes, *The Measurement of Attitudes* (1929). Cada reactivo va acompañado por su valor de escala.

- Creo que la iglesia es actualmente la más grande institución de Estados Unidos. (11)
- Creo en la religión, pero pocas veces voy a la iglesia. (9.6)
- Creo en la sinceridad y la bondad sin ceremonias en la iglesia. (6.7)
- Creo que la iglesia es un obstáculo para la religión porque todavía depende de magias, supersticiones y mitos. (5.4)
- Creo que la iglesia es un parásito de la sociedad. (.2)

Debe quedar claro que el reactivo con un valor de escala de 5.4 tiene un contenido más neutral que cualquiera de las demás

Escalas de Likert

La escala de Likert (Likert, 1932) es fácil de desarrollar y ampliamente utilizada. Aunque su construcción es similar a la de una escala de Thurstone, su desarrollo no requiere tanto tiempo.

He aquí los pasos a seguir para crear una escala de Likert:

- Se escriben afirmaciones que expresan una opinión o un sentimiento acerca de un suceso, objeto o persona. Por ejemplo, si queremos estudiar las actitudes hacia el apoyo del gobierno para las guarderías, los siguientes podrían ser ejemplos de reactivos:
 - El gobierno federal no tiene por qué financiar programas de guardería.
 - El gobierno federal debe dar todo su apoyo a las guarderías.
 - Se debe usar dinero de impuestos para financiar programas de guardería.
- Se seleccionan reactivos que tienen valores positivos y negativos claros (a juicio de quien está creando la escala).
- Se listan las afirmaciones, y a la derecha de cada una se deja un espacio para que el encuestado indique el grado en que está de acuerdo o en desacuerdo, utilizando una escala de cinco puntos como:

TA Totalmente de acuerdo
 A De acuerdo
 1 Indeciso
 D En desacuerdo
 TD Totalmente en desacuerdo

Se pide a los encuestados que encierren en un círculo o marquen su nivel de acuerdo con cada reactivo, como se muestra en la figura 6.4

Instrucciones: Indique en qué medida está de acuerdo o en desacuerdo con cada una de las afirmaciones de la parte inferior encerrando en el círculo uno de los códigos siguientes:

TA significa que usted está *totalmente de acuerdo* con la afirmación
 A significa que usted está de acuerdo con la afirmación
 1 significa que usted está indeciso respecto a la afirmación
 D significa que usted está en desacuerdo con la afirmación
 TD significa que usted está totalmente en desacuerdo con la afirmación

Reactivo	Calificación				
El gobierno federal no tiene por qué financiar programas de guardería.	TA	A ✓	1	D	TD
El gobierno federal debe dar todo su apoyo a las guarderías.	TA	A	1	D	TD ✓
Se debe usar dinero de impuestos para financiar programas de guardería.	TA	A	1	D	TD ✓

Figura 6.4 Conjunto de reactivos de Likert.

Las escalas de Likert se califican asignando un peso a cada punto de la escala, y el puntaje de un individuo es el promedio de todos los reactivos. Sin embargo, la cosa no es tan sencilla. Dado que los reactivos pueden "invertirse" (como en los casos en que se expresan como un negativo, por ejemplo, "No me gusta la escuela"), es preciso ser consistentes e invertir la escala al calificar tales reactivos. La regla es que los reactivos favorables ("Me gusta la escuela") se califican del 5 al 1, donde 5 representa Totalmente de acuerdo. Los reactivos desfavorables ("No me gusta la escuela") se califican del 1 al 5, donde 1 representa Totalmente de acuerdo. De este modo, las calificaciones más altas siempre reflejan actitudes más positivas.

En el ejemplo de la figura 6.4, el primer reactivo está redactado como una expresión negativa, y los dos últimos, como expresiones positivas. Dadas las respuestas que se indican en la figura 6.4, el puntaje para estos tres reactivos sería

Reactivo	Reactivo	Puntaje
1	A	2
2	TD	1
3	TD	1

para dar un puntaje final de $(2 + 1 + 1) / 3 = 1.3$, lo que indica un nivel relativamente intenso de desacuerdo general. Recuerde que el reactivo 1 puntuó de manera invertida porque se expresó como un negativo. Dado que las calificaciones se suman, la creación de una escala de Likert también se conoce como método de calificaciones sumadas.

Pruebas de personalidad

Las pruebas de personalidad evalúan los patrones individuales de conducta estables y son el tipo de prueba más común que se lista en el

anuario Buros Mental Measurement Yearbook (Conoiey & Kramei; 1989). Si bien las pruebas de personalidad pueden ser herramientas de evaluación muy valiosas, su creación es laboriosa y se requiere capacitación para administrarlas, calificarlas e interpretar los puntajes.

Existen básicamente dos tipos de pruebas de personalidad, proyectivas y estructuradas.

Las pruebas proyectivas presentan al encuestado un estímulo un tanto ambiguo y luego piden a la persona formular algún tipo de respuesta. El supuesto en que se basan estos tipos de pruebas es que la persona que está presentando la prueba proyectará (o impondrá) su propia visión del mundo sobre los estímulos y que tales respuestas formarán un patrón que la persona capacitada que administra la prueba podrá evaluar.

La calificación de estos tipos de pruebas y el llegar a conclusiones acerca de los patrones de personalidad y conducta no son cosa del otro mundo. Los psicólogos saben que ciertos tipos de personalidades responden según patrones característicos. Sin embargo, reconocer tales patrones requiere una buena cantidad de tiempo, capacitación y práctica. Ejemplos de estas pruebas son la Prueba de Apercepción Temática y la Prueba Rorschach.

Las pruebas estructuradas utilizan un formato con el cual es probable que usted esté familiarizado, como verdadero-falso, opción múltiple o si-no. En estas pruebas se pide a los sujetos indicar su acuerdo o desacuerdo con un reactivo que describe sus sentimientos hacia ellos mismos (como "Me agrado"). Ejemplos de tales pruebas son el Cuestionario de Dieciséis Factores de Personalidad y el Inventario de Personalidad Multifásico Minnesota. Una de las principales ventajas de las pruebas estructuradas sobre las proyectivas es que las primeras son objetivas en el diseño de los reactivos y son fáciles de calificar. De hecho, las casas editoras de éstas (y de muchos otros tipos de prueba también) ofrecen servicios de calificación. (Sería interesante que averiguara si en su localidad está disponible tal servicio o alguno semejante.) Sin embargo, el hecho de que la prueba sea fácil de calificar nada tiene que ver con la interpretación de los resultados de la prueba. Créame, es mejor dejar las pruebas de personalidad para la "gente grande" que tiene las habilidades y ha recibido la capacitación. De hecho, muchos editores de pruebas de personalidad no le venderán los materiales si usted no puede demostrar que es una persona capacitada (digamos, que tiene un doctorado) o que tiene el aval de una persona capacitada (digamos, su asesor de tesis).

Técnicas de observación

Es posible que el tipo de resultados de pruebas con el que usted está más familiarizado sea el que se obtiene cuando un niño o un adulto presenta una prueba. Ese tipo de prueba convierte al encuestado en el agente activo del proceso de medición. En una clase totalmente distinta de métodos de evaluación del comportamiento, el investigador; que podría ser usted, se convierte en el agente activo. Estos se denominan métodos o técnicas de observación, y son aquéllos en los que el investigador se sitúa filera de la conducta que se está observando y crea una bitácora, notas, o un registro en audio o video de la conducta.

Se emplean muchos términos para describir las actividades de observación, y varios de ellos se han tomado del trabajo realizado por antropólogos y etnólogos como trabajo de campo o observación naturalista. El punto más importante que debemos recordar acerca de los métodos de observación es la razón por la que han sido tan útiles para los científicos de otras disciplinas; es decir; su objetivo primario es registrar el comportamiento sin interferirlo. Como observador; usted deberá hacer todo lo posible por mantenerse al margen de la conducta que está observando para no estorbar ni interferir.

Por ejemplo, si le interesa estudiar el comportamiento de juegos entre niños con y sin discapacidades, bien haría en observarlos desde lejos en lugar de convertirse en parte de su entorno. Estar muy cerca de ellos mientras juegan sin duda tendría un impacto sobre su comportamiento.

Técnicas para registrar el comportamiento

Hay varias técnicas distintas que pueden servir para observar y registrar conductas en el campo, y pertenecen a cuatro categorías generales.

La primera categoría es el registro de duración. Aquí, el investigador emplea un dispositivo para tomar el tiempo y mide la duración de una conducta. Por ejemplo, al investigador podría interesarle conocer qué tanta actividad física ocurre durante el recreo matutino en un jardín de niños. Él o ella podría usar un cronómetro para registrar el tiempo durante el cual un niño realiza actividad física, luego haría lo mismo con otro niño, y así. El investigador está registrando la duración de un suceso específico.

La segunda categoría principal de técnicas de observación del comportamiento es el registro de frecuencia, donde se toma nota de la incidencia o frecuencia de la ocurrencia de una conducta determinada. Por ejemplo, un investigador podría querer registrar el número de veces que un comprador toca y siente la textura de una tela para confeccionar ropa, o el número de comentarios que hace acerca de cierta marca de jabón.

La siguiente categoría de la lista es el registro de intervalo o muestreo de tiempo, donde se observa un sujeto específico durante cierto intervalo de tiempo. Por ejemplo, el investigador podría observar cada uno de los niños de un grupo de juego durante 15 segundos, registrar las conductas objetivo, y luego pasar al siguiente niño para sus 15 segundos. Aquí, el intervalo se refiere al tiempo que el observador se enfoca a un sujeto específico, sin importar lo que el sujeto esté haciendo.

Por último, el registro continuo es aquel en el que se registran todas las conductas del sujeto objetivo sin importar la especificidad de su contenido. Es común que la gente que lleva a cabo estudios de casos observe un niño durante cierto tiempo sin buscar una serie de conductas previamente designadas. Más bien, las conductas que se registran son las que ocurren en el flujo normal de sucesos. Esta es una forma rica y fructífera de recabar información pero tiene la importante desventaja de que la poca planificación que se invierte en el registro de la información exige un procesamiento intensivo de los registros en el momento del análisis.

En la tabla 6.4 se presenta un resumen de estos cuatro tipos de técnicas, lo que cada uno hace y un ejemplo de cada uno.

¿Por qué podrían ser atractivos estos tipos de técnicas? La razón primordial es que el acto mismo de observar algunas conductas interfiere las conductas reales que los investigadores podrían querer estudiar. Por ejemplo, ¿ha entrado alguna vez en un salón de una escuela primaria y ha notado cómo todos los niños voltean a verlo? Algunos niños incluso montan un pequeño espectáculo para usted. Tarde o temprano ese tipo de comportamiento por parte de los niños irá desapareciendo, pero ciertamente usted no obtendrá una visión carente de influencias de lo que sucede ahí. La palabra clave, entonces, es discreción: observar el comportamiento sin alterar la naturaleza de lo que se está observando.

Técnica	Cómo funciona	Ejemplo
Registro de duración	El investigador registra durante cuánto tiempo ocurre una Conducta.	¿Cuánto tiempo posan en interacción verbal dos niños?
Registro de frecuencia	El investigador registra el número de veces que ocurre una conducta.	¿Cuántas veces se hacen preguntas?
Registro de intervalo	El investigador observa un sujeto durante un tiempo fijo.	Dentro de un periodo de 60 segundos, ¿cuántas veces hablan con otra persona los miembros del grupo?
Registro continuo	El investigador registra todo lo que sucede	Durante un periodo de una hora se registra todo el comportamiento de un niño de seis años.

Tabla 6.4 Cuatro formas distintas de observar conductas.

El empleo de cuatro técnicas distintas se ha facilitado de manera considerable con la introducción y disponibilidad de tecnología fácil de usar. Por ejemplo, no hay que sentarse y observar continuamente un grupo de adultos que toman una decisión cuando se puede grabar en video el grupo y luego hacer un análisis a fondo de sus conductas. Asimismo, en lugar de usar lápiz y papel para registrar la conducta cada 10 segundos, usted podría hacer que su computadora personal emitiera un "bip" cada 10 segundos y usted sólo tendría que pulsar una tecla para especificar la categoría de la conducta.

La única reserva que tengo acerca del papel cada vez más importante de la tecnología (que facilita mucho y hace mucho más confiable la recolección de datos) es que el estudiante de técnicas de investigación tal vez nunca tenga la experiencia de usar técnica para reunir información. El muestreo por tiempo, por ejemplo, es una técnica muy útil que también es una buena experiencia. Después podrá usted usar su lujosa computadora personal para continuar.

Toda recolección de datos de este tipo debe efectuarse poniendo especial atención en cuestiones como el anonimato y el respeto por la persona observada. (ya tocamos muchos de estos puntos en el capítulo 2.) Por ejemplo, hay que escoger muy bien dónde y cómo se va a observar. Si bien podría ser muy interesante escuchar la conversación privada de los adolescentes en el baño, esto ciertamente violaría su derecho a la intimidad. Grabar conversaciones telefónicas podría ser una forma efectiva de asegurar el anonimato, porque tal vez usted desconozca el nombre de quien llama (si usted solicita llamadas), pero es preciso enterar a las personas de que sus palabras están siendo grabadas.

¿Técnicas de observación? Tenga cuidado

Ninguna técnica para evaluar conductas es perfecta, y todas están cargadas de problemas en potencia que podrían hundir las mejores empresas. He aquí algunos problemas específicos que conviene considerar si se desea utilizar técnicas de observación:

- Su mera presencia podría afectar el comportamiento que está observando.
- Su propia predisposición o perspectiva podría afectar la forma en que registra el comportamiento, desde lo que decide registrar hasta la forma en que lo registra.
- Usted podría cansarse o aburrirse y pasar por alto aspectos importantes de la conducta que está observando, o ni siquiera percibir la conducta.
- Usted podría modificar la definición de las conductas que desea observar, de modo que lo que se definió como agresión en el tiempo 1 (tocar sin permiso) se redefine en el tiempo 2 porque usted se da cuenta de que no todos los contactos (aunque sea sin permiso) son necesariamente agresivos.

Cuestionarios

Los cuestionarios son un conjunto de preguntas estructuradas y enfocadas que se contestan con lápiz y papel.

Los cuestionarios ahorran tiempo porque permiten a los individuos llenarlos sin ayuda ni intervención directa del investigador (muchos son auto-administrados). De hecho, cuando no es posible estar con los participantes directamente, un cuestionario por correo puede producir los datos necesarios.

Los cuestionarios tienen otras ventajas además de poder auto-administrarse:

- Dado que se usa el correo, es posible estudiar un área geográfica extensa.
- Son más económicos (incluso con lo que cuesta ahora el correo) que las entrevistas personales.
- Es posible que la gente esté más dispuesta a decir la verdad porque su anonimato está prácticamente garantizado.

La objetividad de los datos también hace que sea más fácil compartirlos con otros investigadores y utilizarlos en análisis adicionales. Aunque es posible que haya pasado tiempo desde que se recabaron los datos, puede ser que den pie a preguntas adicionales que estén esperando ser respondidas.

Por ejemplo, en un estudio (Hanson y Ginsburg, 1988), los investigadores utilizaron los resultados de las encuestas "Educación Media y Más Allá" recabados originalmente en la primavera de 1980 de más de 30 000 estudiantes de segundo año de educación media. A estos investigadores les interesaba examinar las relaciones entre los valores de los estudiantes de educación media, sus puntajes en pruebas, calificaciones, problemas de disciplina y status de abandono de estudios. Con una tasa de respuesta original de 84%, estas encuestas produjeron una base de datos muy grande y completa con la cual trabajar. Es posible que la tasa de respuesta haya sido inusualmente alta porque los estudiantes posiblemente eran parte de un público cautivo. En otras palabras, se les proporcionaron los cuestionarios como parte de las actividades normales de la escuela.

No obstante, hay que tener presente que todas estas ventajas no implican una recomendación de salir a recopilar todos sus datos empleando este método. Una de las grandes desventajas de los cuestionarios es que la tasa de completación y devolución es mucho más baja que si usted hace personalmente las preguntas a cada encuestado potencial mediante una entrevista, técnica que veremos un poco más adelante. Cabe esperar una tasa de participación elevada (hasta 100%) si se visita a la gente en sus hogares y se les hace preguntas, pero podría esperarse una tasa de devolución de 35% con un cuestionario pequeño por correo incluso si usted hace un trabajo excelente.

Qué hace que funcione un cuestionario?

¿Qué es un buen cuestionario? Son varios los aspectos de un cuestionario que hacen que "funcione", o que la tasa de devolución con todos los reactivos (o tantos como es posible) contestados sea alta.

Ahora bien, con toda seguridad usted ha completado cuestionarios en algún momento, sean acerca de sus opiniones políticas o de lo que usted busca en un receptor estereofónico. El que los cuestionarios funcionen o no funcionen depende de varios factores que están bajo su control. Examinemos con brevedad cada uno de estos factores, que se resumen en la tabla 6.5 y se desglosan en tres rubros generales: supuestos básicos en los que se fundamenta el cuestionario, las preguntas en sí y el formato en que se presentan los reactivos.

Supuestos básicos

- El cuestionario no exige cosas poco razonables del encuestado.
- El cuestionario no tiene un "propósito oculto".
- El cuestionario solicita información que se supone tienen los encuestados.

Las preguntas

- El cuestionario contiene preguntas susceptibles de ser respondidas.
- El cuestionario contiene preguntas sencillas.

Formato

- Los reactivos y el cuestionario se presentan en un formato atractivo, profesional y fácil de entender.
- Todas las preguntas y páginas están numeradas claramente.

El cuestionario contiene instrucciones claras y explícitas sobre cómo debe llenarse y devolverse.
Las preguntas son objetivas.
Las preguntas están ordenadas, de las fáciles a las difíciles y de lo general a lo específico.
Se incluyen transiciones de un tema al siguiente.
Se dan ejemplos si es necesario.

Tabla 6.5 Puntos importantes que debemos recordar acerca del diseño de cuestionarios.

Supuestos básicos del cuestionario

En la tabla 6.5 se sintetizan cinco puntos importantes en lo tocante a los supuestos que hacemos al diseñar un cuestionario. Es posible que los posibles encuestados estén en la mejor disposición de ayudarlo, pero usted debe ayudarles a ser la clase de encuestado que usted busca.

1. Usted no quiere pedir a los encuestados que llenen un cuestionario de 40 páginas ni que dediquen tres horas del sábado a hacerlo. Su cuestionario debe diseñarse de tal modo que sus exigencias de tiempo, gasto y esfuerzo sean razonables. También conviene evitar hacer preguntas inapropiadas (demasiado personales) o planteadas de manera equivocada. Cualquier cosa que a usted le parezca ofensiva ofenderá también a sus encuestados potenciales.
2. Su cuestionario debe diseñarse pensando en lograr un objetivo y no en recabar información acerca de un tema relacionado pero implícito. Si a usted le interesan las actitudes raciales, deberá dirigir sus preguntas hacia las actitudes raciales y no hacer preguntas enmarcadas en un contexto diferente relacionado indirectamente con su propósito.
3. Si usted quiere averiguar qué tanto sabe un encuestado acerca de un área, debe tener razones para creer que esa persona tiene conocimientos que compartir. Preguntar a un estudiante universitario de primer semestre, en el primer día de clases, qué opina de los beneficios que ofrece la universidad con toda seguridad no proporcionará datos útiles. En cambio, si hace la pregunta en el último día de clases, probablemente obtendrá una mina de oro de información.
4. Anime a los encuestados diseñando un cuestionario que contenga preguntas interesantes, que haga que los participantes sientan deseos por contestar todas las preguntas, y que los estimule para devolverle el cuestionario lleno. Si no puede hacer interesantes sus preguntas, quizá no tenga los conocimientos o el entusiasmo suficientes sobre el tema, y le convendría escoger otro.

Si puede obtener la información de una fuente distinta de un cuestionario, hágalo. Si una entrevista le proporciona una mejor respuesta y datos más exactos, use una entrevista. Si puede averiguar el promedio de calificaciones de un estudiante por otros medios, es mejor dedicar un poco más de tiempo a hacerlo en lugar de abrumar al encuestado con cuestiones que tienen importancia secundaria para sus propósitos.

¿Y las preguntas?

Las preguntas vienen en todos colores y sabores, y algunas son absolutamente horribles. Por ejemplo,

¿Suele sentir ansiedad respecto a presentar un examen y obtener una baja calificación?

¿Puede usted ver por qué la anterior no es una buena pregunta? Por principio, la conjunción *y* hace que sean dos preguntas en lugar de una, lo que vuelve muy difícil determinar qué es lo que está causando la reacción del encuestado. Diseñar buenas preguntas no es imposible; Sólo se requiere algo de tiempo y práctica.

Primero, asegúrese de que sus preguntas puedan contestarse. No pregunte acerca de la actitud de la gente hacia las luchas políticas en algún país extranjero si nada saben de los asuntos de ese país.

Asimismo, redacte de manera directa cada interrogante. La pregunta

¿No hace nunca trampa en sus exámenes?

es rebuscada, utiliza un doble negativo, con la misma facilidad podría formularse así:

¿Hace usted trampa en sus exámenes?

Esta pregunta es más clara y más fácil de contestar con exactitud.

Por último, tenga en cuenta qué tan deseables son socialmente las preguntas. ¿Hay alguien que contestaría amable y positivamente la pregunta

¿Golpea usted a sus hijos?

Desde luego que no, y la información obtenida de preguntas tan directas podría ser de dudoso valor.

El formato del cuestionario

Como puede ver en la tabla 6.5, podemos aplicar varios criterios al formato de un cuestionario, y todos son tan importantes que no prestar atención a una de ellas podría hundir su proyecto.

Por ejemplo, digamos que usted diseña un maravilloso cuestionario con preguntas bien planteadas que se pueden contestar justo en el tiempo ideal, e incluso llama por teléfono a todos los participantes para ver si tienen dudas. Desafortunadamente, ¡usted olvida darles instrucciones detalladas sobre cómo devolver el cuestionario! O quizá usted incluye instrucciones detalladas para la devolución pero olvida decir cómo deben contestar las preguntas.

- Si su cuestionario no consiste en reactivos o preguntas fáciles de leer (claramente impresas, no apiñadas físicamente, etc.), su estudio no tiene mucho futuro. Los reactivos deben organizarse bien, y las copias del cuestionario deben ser claras. El fotocopiado produce buenos resultados, pero es mejor imprimir la cantidad que necesita en una impresora láser, si cuenta con ella. Si puede, haga que un amigo o amiga que sepa de autoedición por computadora le ayude en aspectos como espacios blancos, proporciones y equilibrio.
- Todas las preguntas y páginas deben estar numeradas claramente (por ejemplo, 1, 2,3, 4...). No use combinaciones complejas como 1-1.2011.4.
- Los buenos cuestionarios contienen instrucciones completas y concisas, que le dicen al encuestado exactamente qué hacer ("complete esta sección") y cómo hacerlo ("encierra en un círculo la respuesta", "marque tantas como sean aplicables"). Estas instrucciones también ofrecen indicaciones explícitas sobre la forma de devolver el cuestionario, e incluyen sobres con porte pagado y los datos preimpresos del destinatario, así como un número telefónico al que se puede llamar para obtener más información si es necesario.
- Admítalo, sus encuestados le están haciendo un favor al contestar el cuestionario. Su objetivo es lograr que el mayor número posible de ellos lo hagan. Una forma de animar a los encuestados a que respondan es demostrar que el trabajo de usted está siendo apoyado por un miembro académico o su asesor, y esto puede hacerlo con una carta de presentación o de portada como la que se mostró en la figura 2.3.
- Usted desea una respuesta lo más honesta posible de sus encuestados y, por tanto, debe tener cuidado de que sus preguntas no los estén guiando para contestar de algún modo específico. Las preguntas deben ser objetivas y directas. Una vez más, tenga cuidado con los enunciados socialmente indeseables.
- Las primeras preguntas deberán servir como "calentamiento" para el encuestado. A principio conviene presentar preguntas relativamente sencillas, inocuas y fáciles de contestar ("¿Cuántos años cumplidos tiene?"), para que el encuestado se sienta cómodo. Luego, a medida que las preguntas avancen, podrían hacerse preguntas más complicadas (y personales). Por ejemplo, muchos cuestionarios comienzan con preguntas relativas a datos demográficos como edad, género, raza, etc.; información que la mayoría de la gente proporciona sin recelo. Preguntas posteriores podrían tratar cuestiones como la actitud hacia los prejuicios, asuntos de religión, y demás.
- Cuando su cuestionario cambie de orientación (o temas), es preciso hacérselo saber a' encuestado. Si hay una serie de preguntas acerca de datos demográficos seguido de una serie de preguntas acerca de relaciones raciales, es necesaria una transición de una a la otra, como

Gracias por contestar estas preguntas acerca de usted mismo. Ahora nos gustaría hacerle unas preguntas acerca de sus experiencias con personas que pertenecen a su mismo grupo étnico y las que pertenecen a otros grupos.

- Por último, esfuércese por diseñar un cuestionario fácil de calificar. Siempre que sea posible, proporcione opciones de respuesta objetivas y cerradas, como

27. ¿A cuánto ascienden sus ingresos anuales?

- a. menos de \$20000
 - b. \$20 000 a \$24 999
 - c. \$25 000 a \$29 999
- en vez de

27. Indique por favor sus ingresos anuales: \$ _____

En el primer ejemplo, usted puede capturar una letra como respuesta que se usará en el análisis subsecuente. En el segundo, es preciso tomar primero el número y colocarlo en alguna categoría, lo que implica un paso adicional.

La importancia de la carta de presentación

Una parte indispensable de cualquier cuestionario es la carta de presentación o de portada que lo acompaña. Este mensaje es importante porque ayuda a preparar la escena para lo que vendrá. La importancia de una buena carta de presentación crece en el caso de cuestionarios que se envían por correo a los encuestados, pues establece el sentido de autoridad y comunica la importancia del proyecto.

En la figura 6.5 se muestra un ejemplo de carta de presentación que se usó en el estudio del comportamiento lúdico de los niños con y sin

discapacidad. Observe que esta carta de presentación (que también actúa como forma de consentimiento, lo cual se explicó en el capítulo 2) tiene ciertas características:

- Está escrita en papel membretado de la universidad, lo que causa una impresión favorable en los encuestados y aumenta la probabilidad de que contesten.
- Tiene fecha reciente, lo que indica que la solicitud tiene cierta vigencia.
- La carta está personalizada: no va dirigida a "Estimado participante" sino a "Estimados Sr. y Sra. Margolis".
- Se indican claramente el propósito del cuestionario y la importancia del estudio.
- Se menciona una estimación del tiempo con el fin de que los encuestados sepan cuando deben devolverlo.
- Se promete claramente la confidencialidad y se indica cómo se mantendrá ésta.
- Se hace que los encuestados formen parte del proyecto al decirles que se les enviará una copia de los resultados cuando el estudio haya terminado. Esto puede ayudar a que los encuestados sientan que son una parte importante del estudio.
- La carta tiene una expresión de agradecimiento clara, separada del resto del texto.
- La carta debe tener la firma del jefe del proyecto y la suya. Sabemos que le gustaría poner sólo su nombre, porque usted es quien va a hacer casi todo el trabajo, pero en esta etapa tan temprana de su carrera esta pequeña ayuda del jefe puede ser muy importante.

En nuestra sociedad encontramos por todos lados pruebas para todo, desde seleccionar personas hasta descartarlas, y su uso se ha convertido en uno de los temas más controvertidos entre los científicos sociales y del comportamiento. En definitiva, las pruebas tienen su sitio, y en este capítulo analizamos diferentes clases de herramientas de medición y cómo pueden utilizarse para evaluar de forma confiable y válida la conducta. No obstante, recuerde que también se requiere una cuidadosa formulación de hipótesis y atención a los detalles en todos los aspectos del proyecto de investigación, si se quiere que el método de medición produzca resultados exactos.

Departamento de Psicología
e Investigación Pedagógica
Stetson-Lawrence University, Routhier, RT 98741
02 de septiembre de 1999

Estimados Sr. y Sra. Margolis:

El Departamento de Psicología e Investigación Pedagógica de la Stetson-Lawrence University apoya la práctica de obtener el consentimiento informado de quien corresponda y proteger a los sujetos humanos que participen en investigaciones. Las pruebas que se realizarán y el cuestionario anexo tienen por objeto ayudar en el estudio del comportamiento lúdico de los niños sin discapacidad y con discapacidad.

Se pedirá a una niña sin discapacidad que participe en un juego con otro niño que tenga una discapacidad, en un cuarto provisto de juguetes y libros, y se grabará en videocinta el comportamiento de la niña. Una sesión durará aproximadamente 25 minutos. Nos interesa estudiar la interacción entre niños que tienen una discapacidad y niños sin discapacidad. Esta información es importante porque nos ayudará a desarrollar métodos que aumenten la eficacia de los esfuerzos por integrar niños con discapacidades en las aulas educativas normales.

Nos gustaría que ustedes, Sr. y Sra. Margolis, autorizaran la participación de su hija Maggie y que completaran y devolvieran este formato y el cuestionario anexo, a la dirección que aparece en este impreso. Tal participación es estrictamente voluntaria. Les aseguramos que su nombre no se mencionará en los hallazgos de la investigación. La información sólo se identificará mediante un código numérico.

De acuerdo con el calendario de trabajo del estudio, les agradeceríamos devolver el cuestionario a más tardar el próximo 17 de septiembre.

Ustedes recibirán los resultados del estudio al término de la investigación.

Agradecemos su atención y apreciamos su interés y cooperación.

Atentamente,

Albert Mouline
Estudiante de posgrado
(913)123-4567

Femand Santini
Jefe del proyecto
(913) 1234567

Firma del padre o tutor

Ejercicios

1. Para la información siguiente acerca de dos puntajes en pruebas de aprovechamiento, calcule los índices de dificultad y de discriminación. El asterisco corresponde a la respuesta correcta.

	Reactivo 1			
	Opciones			
	*A	B	C	D
27%Superior	28	15	7	20
27%Inferior	6	12	21	21

	Reactivo 2			
	Alternativas			
	*A	B	C	D
27%Superior	10	7	28	15
27%Inferior	15	0	15	30

2. Escriba un cuestionario con 10 reactivos (tipo Likert) que mida la actitud hacia el hurto. Asegurese de utilizar enunciados tanto positivos como negativos y de que todos los reactivos estén redactados de forma sencilla y se puedan contestar fácilmente. No olvide incluir un conjunto de instrucciones.
3. Consulte, si le es posible, la más reciente edición de Mental Measurements Yearbook de Buros y resuma la reseña de cualquier prueba que se mencione ahí. ¿Qué propósito tiene la prueba? ¿Es positiva o negativa la reseña? ¿Cómo puede mejorarse la prueba?
4. Compare y contraste el uso de cuestionarios y entrevistas como técnicas de medición.
5. Interprete los índices de discriminación y dificultad siguientes:
 - a. $D=.50$
 $d = -.90$
 - b. $D=.90$
 $d = .25$

6. Describa tres características básicas de un cuestionario.

7. Mencione tres ventajas del uso de cuestionarios.

CAPITULO SIETE

Recopilación de datos y estadística descriptiva

En Neil J. Salkind "Métodos de Investigación", Prentice Hall, México, 1998.

Lo que aprenderá en este capítulo

- Cómo iniciar la recopilación de datos
- Cómo iniciar la codificación de sus datos
- Todo acerca de la construcción de un formato de recopilación de datos
- El uso y la importancia de la estadística descriptiva
- La diferencia entre estadística descriptiva e inferencial
- Qué es una distribución de puntajes, y qué diferencias presentan las distribuciones
- Empleo de medidas de tendencia central y variabilidad para describir un conjunto de puntajes
- Cómo calcular la media, la mediana y la moda, y para qué sirven
- Cómo calcular el intervalo, la desviación estándar y la varianza, y para qué sirven
- Qué es la curva normal, y por qué es importante para el proceso de investigación

En todo tipo de labor de investigación, sea un examen histórico del papel de los fármacos para tratar las enfermedades mentales o los efectos de usar un ratón de computadora sobre la coordinación ojo-mano de los niños, es preciso recopilar datos acerca del tema y analizarlos para probar la viabilidad de la hipótesis. Podemos especular hasta hartarnos acerca de la relación entre ciertas variables o acerca de cómo una podría afectar a otra, pero en tanto no haya pruebas objetivas que apoyen nuestras aseveraciones, nuestro trabajo no será más exacto que sacar de un sombrero una de diez posibles respuestas al azar.

En la parte principal de este capítulo explicaremos la recopilación de datos, comenzando con el diseño de los formatos para recabar información y terminando con un análisis del proceso en sí. Una vez que esté familiarizado con estos importantes primeros pasos, pasaremos a una introducción al uso de la estadística descriptiva: un conjunto de herramientas que permiten entender los datos recopilados. Seguiremos hablando del análisis de datos en el siguiente capítulo, "Estadística inferencial". Luego usted podrá aprender a usar su computadora personal y aplicaciones de software como el Paquete Estadístico para las Ciencias Sociales (SPSS) para realizar análisis de datos.

Pasemos pues a la recopilación de datos y al análisis descriptivo.

Preparación para la recopilación de datos

Después de pensar y pensar, de ir a la biblioteca y de formular lo que usted y su asesor creen que es una pregunta importante e interesante (no lo olvide), ha llegado el momento de iniciar el proceso de recopilar sus datos.

El proceso de recopilación de datos implica cuatro pasos:

- la construcción de formatos para recabar la información que servirán para organizar los datos recolectados,
- la **codificación** que sirve para representar esos datos en un formato de recopilación de datos en la forma más eficiente posible,
- la recopilación en sí de los datos, y
- su asentamiento en el formato de recopilación de datos.

Una vez efectuado todo esto, usted estará listo para comenzar a analizar su información.

Durante todo este capítulo utilizaremos un conjunto de datos de muestra que representa 200 juegos de puntajes obtenidos al aplicar pruebas a niños de escuela primaria y secundaria como parte del Programa de Pruebas de Competencia Mínima de Kansas financiado por el Departamento de Educación del Estado de Kansas, Estados Unidos. Estas pruebas de lectura y matemáticas se administran a los niños de los años escolares 2, 4, 6, 8 y 10 en todo ese estado. Cerca de 200000 niños presentan la prueba cada año. Esta muestra en particular consiste en 200 niños, 95 hombres y 105 mujeres, y los datos correspondientes se muestran en el apéndice B. Su profesor puede obtener el conjunto directamente del autor de este libro (¡que soy yo!). A medida que realicemos procedimientos estadísticos sencillos y específicos, utilice algunos de estos datos y realice también los cálculos. Pruébelo, le gustará.

He aquí una lista de la información recopilada para cada niño.

- número de identificación
- género
- año escolar
- número de edificio
- puntaje en lectura
- puntaje en matemáticas

Se obtuvieron seis datos puntuales para cada niño. En la figura 7.1 se muestra una forma de organizar los datos utilizando parte de la información demográfica básica (año escolar y género). La información de la figura 7.1 se generó utilizando SPSS y se imprimió tal como se ve usando ese programa. Usted aprenderá todo acerca del uso de SPSS en el apéndice A.

	GÉNERO	
	1	2
	AÑO ESCOLAR	AÑO ESCOLAR
	Cuenta	Cuenta
2	20	19
4	16	21
6	17	31
8	23	18
10	19	16

Figura 7.1 Frecuencias de año escolar y género para la muestra de prueba de competencia.

El proceso de recopilación de datos

Ahora que (¡espero!) usted tiene su idea bien definida (y sus planes han sido aprobados por su profesor y el comité), es hora de comenzar a pensar en el proceso de recopilar los datos. Esto implica todo desde ponerse en contacto con los sujetos y organizar los viajes de recopilación de datos hasta el registro en sí de los datos en algún tipo de formato que le ayude a ordenar esta información y facilite el proceso de análisis de los datos.

Construcción de formas de recopilación de datos

Una vez que sepa qué información va a recabar y dónde la va a obtener (una parte crítica de su investigación), el siguiente paso es idear un esquema de organización para reunirlos de modo que pueda aplicar fácilmente algunas técnicas que le permitirán analizar y entender sus hallazgos.

Piense en sus **datos brutos** (esto es, desorganizados) como si fueran las piezas de un rompecabezas, y en los resultados de su análisis de datos como si fueran la estrategia a usarse para armar las piezas. La primera vez que abre la caja de un rompecabezas (¿alguna vez ha armado un rompecabezas de una pizza?), las piezas parecen una maraña de colores y formas, y es lo que son. Son los datos brutos. La estrategia que usted usa para armarlos (como "todas las piezas con queso en este montón") son como las herramientas que usamos aquí para analizar los datos.

Cuando los investigadores reúnen datos, su primer paso es crear un **formato de recopilación de datos**. Por ejemplo, la tabla 7.1 es un formato de recopilación de datos que podría servir para registrar los puntajes y otra información una vez que se han calificado las pruebas. Observe que se incluyen en el formato los posibles valores (si se conocen) de todas las variables, a fin de facilitar el resto. Por ejemplo, los niños se codificarán como 1 y las niñas como 2, etc. Hablaremos más de la codificación en breve.

Un criterio que podemos usar para juzgar si un formato de recopilación de datos es clara y fácil de usar es mostrársela a alguien (digamos, otro estudiante de su grupo) que no esté familiarizado con su proyecto. Luego pídale a esa persona que tome datos de la fuente

Número de identificación	Género 1 = Masculino 2 = Femenino	Año escolar (2,4,6,8,10)	Edificio (1 al 6)	Puntaje en lectura	Puntaje en matemáticas
001	2	8	1	55	60
002	2	2	6	41	44
003	1	8	6	39	37
004	2	4	6	56	59
005	2	10	6	32	32

Tabla 7.1 los primeros cinco de 200 casos en el conjunto de puntales de competencia.

de datos primaria (como la prueba de lectura misma) y los introduzca en el formato de recopilación de datos. ¿Esa persona sabe qué hacer y cómo hacerlo? ¿Es obvio qué va en cada lugar? ¿Queda claro qué significan las entradas? Todas estas preguntas deberán contestarse con un categórico "sí".

La clave para diseñar un formato de recopilación de datos efectivo es la cantidad de planificación que usted invierte en el proceso. Usted podría usar el formato de prueba mismo como formato de recopilación de datos si toda la información que necesita se registra de tal manera que está fácilmente accesible para el análisis de datos. Tal vez en la parte superior del cuadernillo o cuestionario de la prueba hay espacios para registrar toda la información pertinente además de los resultados de la prueba. Esto significa que usted no tendrá que andar buscando todos los datos; podrá hallarlos hasta arriba en la primera página. Un plan así reduce la posibilidad de errores en la transferencia de los datos originales al programa estadístico que se usa para analizar los datos.

En la tabla 7.1 se muestran los primeros siete casos registrados en la forma de recopilación de datos ya llena. Las columnas se organizan por variables, y la información acerca de cada estudiante se introduce como fila individual. Estos cinco casos son los primeros de los 200 (del apéndice B) que se usarán en secciones posteriores de este capítulo para demostrar diversas técnicas de análisis de datos.

Recuerde, el formato de datos que construya deberá ser fácil de entender y usar, porque es el vínculo principal entre los datos originales y el primer paso del análisis de datos. Muchos investigadores ponen a dos personas a trabajar en la transferencia de datos de la hoja original al formato de datos para minimizar el número de posibles errores. Ésta es una razón por la que a veces se usa papel milimétrico o algún otro formato que incluye líneas horizontales y verticales, como se ve en la tabla 7.1.

He aquí algunas sugerencias generales para construir un formato de recopilación de datos. Cada una de ellas está ilustrada en la tabla 7.1.

- Utilice una línea para cada sujeto. Si su formato de datos necesita mucho espacio, hasta podría usar una página por sujeto.
- Utilice papel con columnas o cuadrículas como el papel milimétrico.
- Registre los números de identificación de los sujetos como filas y los puntajes u otras variables como columnas.

- Incluya espacio para toda la información que desea registrar además de información que piensa que tal vez podría registrar en el futuro. Por ejemplo, si está realizando un estudio en el que habrá un seguimiento, deje espacio para las series de puntajes que se introducirán después.
- Conforme vaya llenando los formatos de recopilación de datos, prepare copias y guárdelas en otro lugar por si acaso se pierden los datos originales o su otro registro de recopilación de datos.
- Tal vez convenga fechar cada formato y marcarlo con sus iniciales en el momento de llenarlo.

Recopilación de datos con lectores ópticos

Si recopilará datos en los que las respuestas del sujeto se registran como una de varias opciones (como en las pruebas de opción múltiple), tal vez le convenga calificar los resultados de la prueba utilizando una **hoja para lector óptico** que se califica con un **lector óptico**. Es probable que usted ya haya presentado pruebas en este tipo de hojas.

Las respuestas de estas hojas especiales se leen con un lector óptico y cada respuesta se compara con una clave (que es otra hoja que usted prepara). Entonces, el lector registra las respuestas correctas e incorrectas y marca un puntaje final en la parte superior de la hoja. Ya puede usted imaginarse los beneficios que semejante técnica puede ofrecer:

- Es muy rápida. La calificación manual de 50 pruebas de opción múltiple con 100 reactivos cada una, fácilmente puede tomar horas.
- Estos lectores son más exactos de lo que cualquiera de nosotros puede ser; casi nunca cometen errores.
- Las respuestas leídas ópticamente hacen posible un análisis adicional de los reactivos individuales de la prueba, como los índices de dificultad y discriminación de los que hablamos en el capítulo 6.

¿Son caras estas máquinas? Su precio de unos \$20000 dólares podría ser demasiado para su presupuesto, pero el tiempo y el dinero que ahorran bien podría cubrir el costo. Imagine tener todas las pruebas calificadas el mismo día que usted termina de administrarlas.

Por tanto, si puede, utilice hojas para lector óptico o, si resulta apropiado, transfiera los datos originales a una de estas hojas para facilitar su trabajo y hacerlo más preciso. Casi todas las universidades importantes cuentan con lectores ópticos. Además, varias compañías publican pruebas diseñadas para usar hojas de respuesta especiales que luego se devuelven a la compañía para que las califique.

Cabe hacer una pequeña advertencia. El hecho de que ésta sea una metodología atractiva y pueda ahorrarle tiempo no debe hacerlo caer en la trampa de creer que la hoja para lector óptico es la única forma de recopilar y calificar datos. Si lo hace, terminará tratando de ajustar sus objetivos a un marco de evaluación que podría no ser apropiado para la pregunta que usted está haciendo.

Codificación de datos

Los datos se codifican cuando se transfieren del formato de recopilación original (como el cuadernillo de prueba) a un formato que se presta para el análisis de datos.

Por ejemplo, un niño podría ser de sexo masculino o femenino. Las letras que constituyen las palabras *masculino* o *femenino* no se introducirían en el formato de datos real; más bien, el sexo del niño se codificaría con el valor 1 para representar sexo masculino y 2 para representar sexo femenino, como vimos en la figura 7.1 y en la tabla 7.1. En este ejemplo, el género se codifica como un 1 o un 2. Asimismo, el grupo étnico o cualquier otra variable *categorica* se puede introducir como un número de un solo dígito (en tanto haya menos de 10 categorías que utilicen los numerales del 0 al 9). En la tabla 7.2 se muestran varios tipos de datos distintos y cómo se podrían codificar para los puntajes en matemáticas y lectura de la muestra.

El empleo de un solo dígito (en lugar de una palabra) no sólo ahorra espacio y tiempo de captura; cuando llega el momento de analizar los datos también es mucho más fácil ha-

Variable	Intervalo de datos posibles	Ejemplo
Número de identificación	001 al 200	138
Género	102	2
Año escolar	1,2,4,6,8 o 10	4
Edificio	1 al 6	1
Puntaje en lectura	1 al 100	78
Puntaje en matemáticas	1 al 100	69

Tabla 7.2 Codificación de datos.

cer que la computadora busque todas las entradas de sexo que son iguales a 1, en vez de *masculino*.

La única regla para codificar datos es usar códigos lo más sencillos y carentes de ambigüedad que sea posible sin perder el verdadero

significado de los datos mismos. Por ejemplo, es perfectamente aceptable codificar un muchacho de cuarto año como un 4 para el año escolar y un 1 para el sexo, pero no es recomendable usar letras (como M y F) porque es más difícil trabajar con ellas. Además, no combine categorías; por ejemplo, no use 41 (en vez de 4y 1) para indicar que un sujeto está en cuarto año (4) y es hombre (1). El problema aquí es que usted no podrá separar después el año escolar y el sexo como factores, y sus datos perderán gran parte de su valor.

Lo importante aquí es siempre registrar los datos como elementos lo más explícitos y discretos que sea posible. Esto siempre le permitirá combinar criterios para los datos después, durante el proceso de análisis. No lo haga desde el principio.

Los diez mandamientos de la recopilación de datos

No deje que nadie lo engañe. El proceso de recopilación de datos puede ser largo y arduo, incluso si sólo implica administrar un sencillo cuestionario de una página a un grupo de padres de familia. Probablemente será la parte de su proyecto que más tiempo consuma. Si va a realizar una investigación histórica, es probable que pase casi todo su tiempo en la biblioteca buscando en libros y publicaciones periódicas, o tal vez entrevistando a personas que son importantes para su tesis. Si va a recopilar datos empíricos, tendrá que organizarse de otra manera.

Pues bien, aquí los tenemos: los diez mandamientos para asegurar que obtendrá los datos que necesita. No los grave en piedra como los Diez Mandamientos originales, pero si los obedece podrá evitar errores que podrían ser fatales.

Primero, cuando comience a considerar un proceso de investigación, *piense desde entonces en el tipo de datos que tendrá que recopilar para contestar su pregunta.*

Segundo, mientras *piensa en el tipo de datos que recabará, piense también de dónde los va a obtener.* Si va a usar la biblioteca para consultar datos históricos o a acceder a archivos de datos que ya se recopilaron (¡una estrategia muy cómoda!), tendrá pocos problemas de logística. Pero, ¿y si desea evaluar, por ejemplo, la interacción entre los niños recién nacidos y sus padres? ¿O la actitud de los maestros hacia formar un sindicato? ¿O la edad a la que las personas de más de 50 años piensan que están viejas? Todas estas preguntas requieren obtener respuestas de la gente, y a veces no es fácil encontrar a esa gente. Comience desde ahora.

Tercero, *asegúrese de que el formato de recopilación de datos que esta usando sea claro y fácil de usar.* Practique con un conjunto de datos piloto o artificiales para asegurarse de que sea fácil pasar de las hojas de puntaje originales al formato de recopilación de datos.

Cuarto, una vez que transfiera puntajes a su formato de recopilación de datos, *prepare una copia del archivo de datos y guárdela en un lugar distinto.* Esta regla no implica que sea preciso duplicar el instrumento de recopilación de datos original para cada participante, sea un cuadernillo de prueba de aprovechamiento o un juego de dibujos de figuras. Más bien, una vez que haya terminado de calificar y haya transferido la información a las hojas de recopilación de datos, guarde una copia de dichas hojas en un lugar aparte. Si registra sus datos en forma de archivo de computadora, como una hoja de cálculo (de lo cual hablaremos más adelante), ¡asegurese de hacer un respaldo! Recuerde, hay dos tipos de personas: las que perdieron sus datos y las que los van a perder.

Quinto, *no dependa de otras personas para que recopilen o transfieran sus datos* si no los ha capacitado personalmente y tiene confianza en que entienden el proceso de recopilación de datos tan bien como usted mismo. Es magnífico tener gente que lo ayude a uno, y ayuda a conservar el ánimo durante esas largas sesiones de recopilación de datos, pero a menos que sus ayudantes posean una competencia más allá de toda duda, fácilmente podrían echar por tierra toda su ardua labor y planificación.

Sexto, *haga un programa detallado de cuándo y dónde recabará sus datos.* Si es necesario visitar tres escuelas, y en cada una es preciso aplicar la prueba a 50 niños en forma individual durante 10 minutos cada uno, esto implica 25 horas de pruebas. Eso no significa que baste con destinar 25 horas de su programa a esta actividad. ¿Y el tiempo de traslado de una escuela a otra? ¿Y el niño que estaba en los sanitarios cuando le tocó su turno, y hay que esperarlo 10 minutos hasta que regresa al salón de clases? ¿Y el día que usted llega y se encuentra con que un personaje de la televisión va a visitar la escuela? Los posibles contratiempos son infinitos. Prepárese para cualquier cosa y asigne entre 25 y 50% más de tiempo en su programa por sucesos imprevistos.

Séptimo, lo antes posible, *cultive las posibles fuentes de su grupo de sujetos.* Puesto que usted ya tiene ciertos conocimientos en su propia disciplina, probablemente conoce personas que trabajan con el tipo de población que usted busca o que podrían ayudarle a obtener acceso a esas muestras. Si usted pertenece a una comunidad universitaria, es probable que haya cientos de personas más compitiendo por la misma muestra de sujetos que usted necesita. En vez de competir, ¿por qué no buscar un distrito escolar o grupo social u organización cívica u hospital más alejado (digamos, a 30 minutos), donde podría obtener una muestra sin tantos competidores?

Octavo, trate de *ponerse en contacto otra vez con los sujetos que faltaron a su sesión de pruebas o entrevista.* Llámelos de nuevo y trate de hacer otra cita. Si adquiere el hábito de pasar por alto posibles participantes, se le hará cada vez más fácil reducir la muestra hasta un tamaño excesivamente pequeño.

Noveno, *nunca deseche los datos originales,* como los cuadernillos de prueba, las notas de entrevistas, etc. Otros investigadores podrían querer usar la misma base de datos, o usted podría tener que regresar a los materiales originales para obtener más información.

¿Y el número 10? Obedezca los otros nueve. ¡Es en serio!

Preparación para el análisis de datos

Muy bien. Ya pasó usted muchas largas y arduas horas preparando una propuesta que valga la pena, un formato de recopilación de datos útil, y acaba de invertir seis meses en recabar la información y capturarla en un formato susceptible de análisis. ¿Qué sigue en la lista?

Primero, utilizando **estadística descriptiva**, usted puede describir algunas de las características de la *distribución de puntajes* que recopiló, como el puntaje promedio para una variable o el grado en que un puntaje varía respecto a otro. Por último, una vez que los datos estén organizados de una forma que permita examinarlos de cerca, aplicará el conjunto de herramientas llamado **estadística inferencial** para tomar decisiones acerca de la relación que hay entre los datos que recabó y sus hipótesis originales, y qué tan generalizables son estos resultados a un número mayor de sujetos que los que se sometieron a las pruebas.

El resto de este capítulo se ocupa de la estadística descriptiva. El capítulo 8 trata sobre la estadística inferencial.

¿Quién iba a pensar que usted iba a estar inscrito en un curso en el que aparece una y otra vez la temida palabra *estadística* (a veces llamada *sadística*)? Pues bien, ahora usted va a aprender algo acerca de esta misteriosa parte del proceso de investigación, y es posible que hasta se encariñe con el conjunto de potentes herramientas que describiremos. En vista de lo prevaeciente que es la ansiedad y la preocupación acerca de esta área del proceso de investigación, he aquí unos cuantos consejos para asegurar que usted no se convierta en miembro del grupo que padece el complejo de "No puedo" antes de siquiera intentarlo:

- Lea el resto del capítulo sin poner mucha atención en los ejemplos. Sólo trate de adquirir una idea general de su organización y del material que se cubre.
- Vuelva al principio de esta sección y siga cuidadosamente cada uno de los ejemplos que se presentan, paso a paso. Si se topa con problemas, repita el paso 1.
- Si las cosas se ponen en verdad difíciles, tómese un descanso y luego regrese a la parte del capítulo o ejercicio que sí entendió claramente. Luego continúe a partir de ese punto.
- Tenga presente que la mayor parte de la estadística consiste en entender y aplicar unos cuantos supuestos sencillos y básicos. La estadística no es una rama avanzada y compleja de las matemáticas.

Resuelva los ejercicios tanto a mano como con una calculadora para asegurarse de que entiende las operaciones básicas que implican. Cuando aprenda a usar SPSS (apéndice A), vuelva a resolver los ejercicios. Cuanto más practique estas técnicas, mayor habilidad adquirirá para usarlas como herramientas para entender sus datos.

Estadística descriptiva

El primer paso del análisis de los datos consiste en describirlos. Esto generalmente implica calcular una serie de medidas de estadística descriptiva, así llamadas porque describen las características generales de un conjunto o *distribución de puntajes*. Efectivamente, estas cifras permiten al investigador (o al lector del informe de la investigación) obtener una primera impresión exacta del aspecto que presentan los datos.

Antes de estudiar las diferentes medidas de estadística descriptiva, veamos qué es realmente una distribución de puntajes y cómo puede ayudarnos a entender mejor los datos.

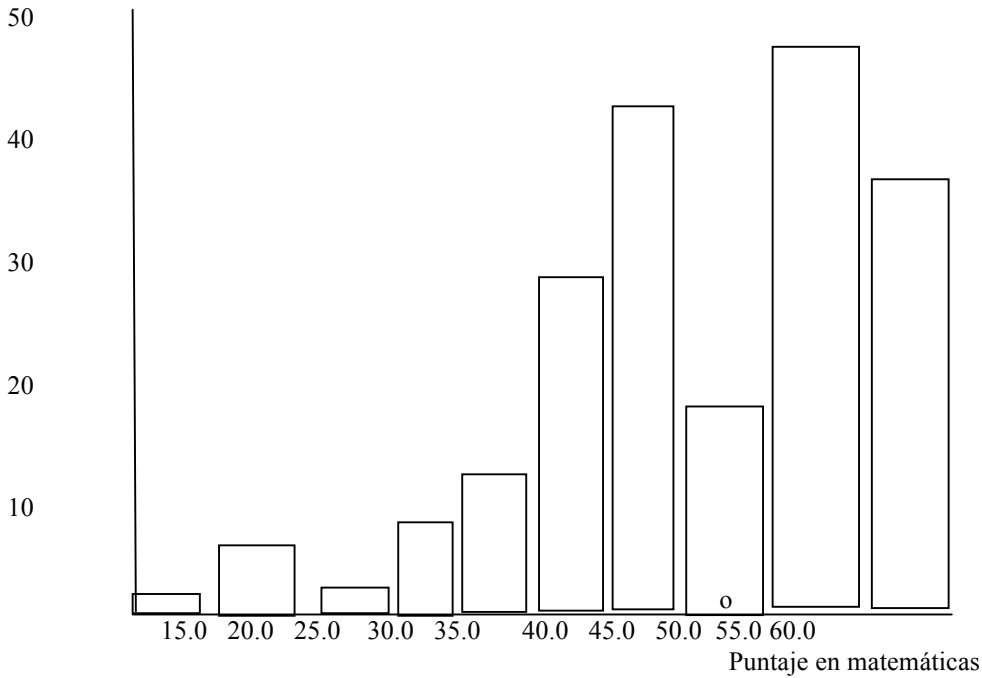
Distribuciones de puntajes

Si usted le pregunta a su mejor amigo o amiga su edad, habrá recabado un elemento de información o un dato puntual para ese individuo. Si usted recaba un elemento de información para más de un individuo, digamos las edades de todos los integrantes de su grupo, tendrá un conjunto de puntajes y varios datos puntuales. Dos o más datos puntuales constituyen una distribución de puntajes. Por ejemplo, en la figura 7.2 vemos una forma de representar una distribución de puntajes utilizando un tipo especial de gráfica (llamada histograma) para representar la distribución de los puntajes en matemáticas obtenidos por 200 niños. Por cierto, esta gráfica se creó con SPSS.

El eje vertical corresponde a la frecuencia con que ocurre un puntaje en particular. El eje horizontal, o eje X, corresponde al valor del puntaje. En esta figura cada banda representa 5 puntos a lo largo de la escala. Por ejemplo, aproximadamente 12 niños tuvieron puntajes entre 32.5 y 37.5. Con base en la forma de la distribución, podemos decir varias Cosas acerca de este conjunto de 200 puntajes con un sencillo examen visual del histograma.

- La mayoría de los niños tienen puntajes en la mitad superior de la distribución.
- El puntaje obtenido por la mayor cantidad de niños es aproximadamente 55.

Puntajes en matemáticas por frecuencia



Desv. Est. = 10.02
Media = 47.4
N=200.00

Figura 7.2 Distribución de los puntajes en matemáticas para todos los niñas, año escolar de 2° a 10°.

Comparación de distribuciones de puntajes

Es posible visualizar más de una distribución de puntajes simultáneamente y compararlas visualmente unas con otras. Por ejemplo, en la figura 7.3 se muestran distribuciones individuales para los puntajes en lectura de los niños de segundo y décimo año en la misma gráfica. Podemos ver que las formas de las curvas son similares, pero están desplazadas tanto como difieren entre sí sus puntos centrales. Ésta es sólo una diferencia en las características de las distribuciones. Veamos cómo pueden compararse las distribuciones.

Medidas de tendencia central

Una propiedad de una distribución de puntajes es que tiene un promedio, es decir un valor individual que es más representativo de esa distribución o conjunto de puntajes. Hay tres tipos de **promedios o medidas de tendencia central**: la media, la mediana y la moda.

Segundo año

Décimo año

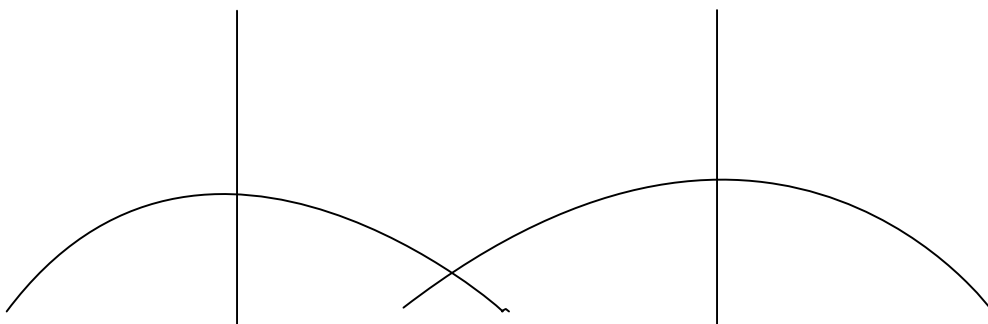


Figura 7.3 Das distribuciones distintas para la misma variable.

La media

La media es la sumatoria de un conjunto de puntajes dividida entre el número de puntajes. Es probable que usted haya calculado muchas medias a lo largo de los años pero les haya llamado promedios, como la cantidad promedio de dinero que usted necesita para cubrir sus gastos o llenar el tanque de gasolina de su automóvil, o el promedio decimal de sus calificaciones durante los últimos tres semestres.

Hay varios tipos de promedios. El que exploraremos aquí es la media aritmética, que es la medida de tendencia central que más se usa. La fórmula de la media es:

$$\bar{X} = \frac{\sum X}{n}$$

donde \bar{X} = la media

\sum = signo de sumatoria

X = puntajes individuales

n = tamaño de la muestra

Para calcular la media, siga estos pasos:

1. Sume todos los puntajes del grupo para obtener un total.
2. Divida el total de todos los puntajes entre el número de observaciones.

Por ejemplo, el puntaje medio en la prueba de lectura para los primeros 10 estudiantes es 47.3. Los primeros 10 puntajes son 55, 41, 46, 56, 45, 46, 58, 41, 50 y 35. Su total es 473, y si lo dividimos entre 10 (el número de observaciones), el resultado es 47.3.

En este ejemplo, 47.3 es el valor que mejor representa la posición más central del conjunto de 10 puntajes. Para los 200 puntajes del apéndice B, la media es de 48.6 en la prueba de lectura y de 47.4 en la de matemáticas. Estos valores se calcularon del mismo modo, sumando los 200 puntajes y dividiendo el total entre el número de puntajes del conjunto, 200.

La media de cualquier variable se puede calcular empleando el mismo método. Sin embargo, como vimos en el capítulo 3, no tiene sentido sumar valores en el nivel nominal (como los que representan un identificador o el sexo) porque el resultado no tiene significado. (¿Qué obtenemos cuando sumamos un niño y una niña y dividimos entre dos?)

La mediana

La mediana es el puntaje de una distribución por arriba del cual se encuentra la mitad de los puntajes. Por ejemplo, en un conjunto de puntajes sencillos como 1, 3 y 5, la mediana es 3. Si se añadiera otro puntaje, digamos 7, la mediana sería el valor que está entre 3 y 5, o sea, 4. Aquí, 50% de los puntajes queda arriba del valor 4 (y, desde luego, 50% queda abajo).

Para calcular la mediana cuando el número de puntajes del conjunto es impar siga estos pasos:

1. Ordene los puntajes del más bajo al más alto.
2. Cuente el número de puntajes.
3. Escoja el puntaje que está en medio como la mediana.

Por ejemplo, aquí vemos cómo se ordenaron los puntajes en lectura de 15 alumnos de segundo año del valor más bajo al más alto. El octavo puntaje (el puntaje que ocupa la octava posición en el grupo) es la mediana. En este caso, ese valor es 43.

31 33 35 38 40 41 42 43 44 46 47 48 49 50 51

Para calcular la mediana si el número (no la sumatoria) de puntajes del conjunto es par, siga estos pasos.

1. Ordene los puntajes del más bajo al más alto.
2. Cuente el número de puntajes.
3. Calcule la media de los dos puntajes que están en medio; ésa es la mediana.

Por ejemplo, los 14 puntajes siguientes se ordenaron del más bajo al más alto. La mediana se calculó sumando los puntajes séptimo y octavo (los puntajes que ocupan las posiciones séptima y octava del grupo, 42 y 43) y dividiendo entre 2. El resultado es 42.5.

31 33 35 38 40 41 42 43 44 46 47 48 49 50

La moda

La moda es el puntaje que ocurre con mayor frecuencia. ¡Cuidado! No es el número de veces que ocurre el puntaje, sino el puntaje mismo. Si tenemos los números siguientes

58 27 24 41 27 26 41 53 14 29 41 53 47 28 56

La moda es 41, no 3 (el número de veces que ocurre el valor 41).

La moda es útil principalmente en el caso de datos nominales como el sexo. En el con-junto de datos de competencia, la moda para el sexo es femenino, ya que hay 105 mujeres y 95 hombres. La moda no es la frecuencia con que ocurre el valor "femenino", que es 105. La moda es una opción excelente si se desea un panorama general de cuál clase o categoría ocurre con mayor frecuencia.

¿Cuál medida cuándo?

La media, la mediana y la moda nos dan diferentes tipos de información y deben usarse de maneras diferentes. En la tabla 7.3 se resumen las ocasiones en que conviene usar cada una de ellas. Como puede verse, el empleo de una u otra medida de tendencia central depende del tipo de datos que se están describiendo.

¿Qué medida de tendencia central debe usarse?		
Medida	Nivel de medición	Ejemplos
Media	Razón e intervalo	rapidez de respuesta, edad en años
Mediana	Ordinal	lugar en el grupo, orden de llegada
Moda	Nominal	afiliación política, color de ojos, etnicidad

Tabla 7.3 Comparación de la media, la mediana y la moda.

Por ejemplo, para describir datos en el nivel de intervalo o de razón (como el número de palabras correctamente escritas) lo mejor es la media, pues proporciona relativamente más información que la mediana o la moda. La regla de oro es: si los datos son apropiados, y se puede, debe usarse la media.

La mediana es más apropiada para datos ordinales u ordenados. Por ejemplo, el conjunto de puntajes 7, 22, 24, 50, 66, 76 y 100 tiene la misma mediana (50) que el conjunto de puntajes 49, 50 y 51, pero las distribuciones son muy diferentes. La mediana también es la opción adecuada cuando se incluyen puntajes extremos en la muestra. Por ejemplo, si tenemos los salarios de cinco empleados: \$21 500, \$27600, \$32000, \$18 750 y \$82000, la mediana es el puntaje de en medio (o el que está en tercer lugar), que es \$27600. La media, en cambio, es \$36 370. Examinemos la gran diferencia entre estos dos valores. ¿Cuál medida cree usted que represente mejor el conjunto de cinco puntajes, y por qué? Si contestó que la mediana, tiene razón. Ciertamente no queremos un "promedio" (\$36 370) que es mayor que el segundo valor más grande del conjunto (\$32000). Este número no sería muy representativo, lo cual es el propósito primordial de cualquier medida de tendencia central. La conclusión que podríamos sacar de este ejemplo es que la mediana es mejor cuando un conjunto de puntajes es extremadamente asimétrico o desbalanceado. Aquí es el dato puntual de \$82000 lo que echa a perder las cosas.

La moda es la mejor opción cuando los datos son de naturaleza cualitativa (nominales o categóricos), como sexo, color del pelo, grupo étnico, escuela o pertenencia a un grupo. Aunque no es común ver que se informe la moda en la bibliografía de investigación (porque podría no tener sentido promediar los valores de variables nominales), es la única medida de tendencia central que se puede usar con información de nivel nominal.

Es obvio que la media nos permite aprovechar el máximo de información (cuando está disponible) y por ello es que suele ser la medida de tendencia central más informativa. Cuando los investigadores pueden hacerlo, escogen variables para las que se puede calcular este tipo de promedio.

Medidas de variabilidad

Acabamos de ver cómo un conjunto de puntajes se puede representar con diferentes tipos de promedios. Pero el promedio no basta para describir plenamente un conjunto de puntajes. Hay otra cualidad o característica importante que describe el grado de variabilidad o dispersión de un conjunto de puntajes.

La **variabilidad** es el grado de dispersión que caracteriza a un grupo de puntajes y es el grado en que un conjunto de puntajes difiere de alguna medida de tendencia central, generalmente la media. Por ejemplo, el conjunto de puntajes 1, 3, 5, 7 y 9 (que tiene una media de 5)

tiene un mayor grado de variabilidad que el conjunto de puntajes 3, 4, 5, 6 y 7, que tiene la misma media. El primer conjunto sencillamente está más "disperso" o "extendido" que el segundo.

Hay varias medidas de la variabilidad, y las veremos una por una.

El intervalo

El intervalo es la diferencia entre el puntaje más alto y el más bajo de una distribución; es la medida más sencilla y directa de qué tan disperso está un conjunto de puntajes.

Por ejemplo, para un conjunto de puntajes como éste:

31 33 35 38 40 40 41 41 41 42 43 44 46 47 48 48 49 49 50 51

el intervalo es 20 (o sea, 51 - 31). En los datos de lectura que se están usando como ejemplo de conjunto de datos grande (demasiado grande para listarlo en orden aquí), el intervalo de los puntajes en matemáticas es 45, o sea, 60-15. El intervalo es una medida burda que indica la dispersión general o el tamaño de la diferencia entre extremos.

La desviación estándar

La desviación estándar es la medida de variabilidad de uso más común. La desviación estándar es la cantidad promedio en que cada uno de los puntajes individuales varía respecto a la media del conjunto de puntajes. Cuanto mayor es la desviación estándar, más variable es el conjunto de puntajes. Si todos los puntajes de una muestra son idénticos, como 10, 10, 10 y 10, no hay variabilidad, y la desviación estándar es cero.

La fórmula para calcular la desviación estándar es

$$s = \sqrt{\frac{\sum(X - \bar{X})^2}{n-1}}$$

- donde s = desviación estándar
 \sum = signo de sumatoria
 \bar{X} = puntaje bruto
 X = media de la distribución
 n = tamaño de la muestra

Para calcular la desviación estándar, siga estos pasos, que se muestran en la tabla 7.4. Calcularemos la desviación estándar para el siguiente conjunto de 10 puntajes:

13 14 15 12 13 14 13 16 15 9

1. Liste todos los puntajes originales y calcule la media (que es 13.4).
2. Reste cada puntaje individual de la media (13.4) y coloque estos valores en la columna titulada "Desviación respecto a la media". Observe que la suma de todas estas desviaciones (alrededor de la media) es 0.

¿Recuerda que la desviación estándar se definió como la cantidad promedio de desviación? Tal vez usted se esté preguntando por qué no paramos aquí, ya que hemos calculado un promedio. La razón es que este promedio siempre es 0 (hablaremos más de esto en un momento). Para deshacernos del valor de cero, elevamos al cuadrado cada desviación.

Puntajes originales	Desviación respecto a la media	Desviación al cuadrado
13	-.04	.26
14	0.6	.36
15	1.6	2.56
12	-1.4	1.96
13	-.04	.16
14	.6	.36
13	-0.4	.16
16	2.6	6.76

	15	1.6	2.56
	9	-4.4	19.36
Sumatoria	134	0	34.4
Promedio	13.4	0	
		Desviación estándar	1.96
		Varianza	3.82

Tabla 7.4 Cálculo de la desviación estándar.

3. Eleve al cuadrado cada una de las desviaciones y colóquelas en la columna rotulada "Desviaciones al cuadrado".
4. Sume las desviaciones al cuadrado (el total deberá ser 34.4).
5. Divida la sumatoria de las desviaciones al cuadrado (34.4) entre el número de observaciones menos 1 (que es 9 en el ejemplo) para obtener 3.82.

Dividimos entre 9 y no entre 10 porque si hay error es preferible que sea en el lado conservador aumentando artificialmente el valor de esta estadística descriptiva. Como se dará cuenta, a medida que aumenta el tamaño de la muestra (digamos, de 10 a 100), el ajuste de restar 1 al denominador cada vez afecta menos la diferencia entre el valor predispuesto (con el tamaño total de la muestra como denominador) y el no predispuesto (con el tamaño de la muestra menos 1 en el denominador).

6. Calcule la raíz cuadrada del valor obtenido en el paso 5 (3.82) para obtener la desviación estándar, que es 1.96. El símbolo para la desviación estándar es s .

¿No se muere de ganas por saber por qué sacamos la raíz cuadrada? Porque nos interesa regresar a los valores originales, y tuvimos que elevarlos al cuadrado en el paso 3 (para deshacernos de las desviaciones negativas, pues de lo contrario su suma siempre sería cero y toda desviación estándar sería cero).

Algunos de los números que se obtienen durante el cálculo de la desviación estándar son muy interesantes. Examinemos la suma de la desviación respecto a la media. ¿Sabe usted por qué (siempre) es cero? Porque la media (de la cual se resta cada uno de los puntajes) representa el punto alrededor del cual la suma de las desviaciones siempre es igual a cero. Si la sumatoria de esta columna no es cero, o bien se calculó incorrectamente la medida o los valores restados son incorrectos.

Otra medida de la variabilidad que a menudo se ve en los informes de investigación es la varianza, que es el cuadrado de la desviación estándar. La varianza tiene una fórmula idéntica a la de la desviación estándar con excepción de la raíz cuadrada. Así como la varianza es el cuadrado de la desviación estándar la raíz cuadrada de la varianza es la desviación estándar. El símbolo de la varianza es s^2 .

Para el conjunto de 200 puntajes de competencia en lectura y matemáticas que hemos usado como ejemplo, la desviación estándar para la lectura es de 7.22, y para matemáticas, de 10.02. La varianza para la lectura es de 52.13, y para matemáticas, de 100.40. Usted podrá confirmar estos valores más adelante en el apéndice A, donde se explica el uso de *SPSS*.

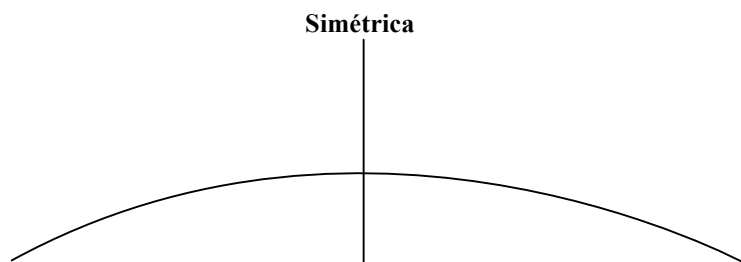
Estudio de las distribuciones

Aunque hemos visto varias medidas de tendencia central y variabilidad, sólo necesitamos dos para darnos una muy buena idea de las cualidades de una distribución: la media y la desviación estándar. Con estas dos medidas de estadística descriptiva podemos entender plenamente la distribución y lo que significa.

La curva normal, o de campana

Ya vio usted la forma de la figura 7.4. Esta curva se denomina comúnmente curva normal o de campana, y es la que representa la forma en que están distribuidas muchas variables (como altura y peso). La curva normal tiene algunas características muy interesantes:

- La media, la mediana y la moda son todas el mismo valor (representado por el punto en que la línea vertical cruza el eje X).
- La curva es simétrica alrededor de su punto medio, lo que implica que las mitades izquierda y derecha de la curva son imágenes especulares.



Eje x

Media
Mediana
Moda

Figura 7.4 Curva normal a de campana.

- Las colas de la curva se acercan más y más al eje X, pero nunca lo tocan; es decir; la curva es *asintótica*.

De hecho, muchas medidas de estadística inferencial (que veremos en el siguiente capítulo) se basan en el supuesto de que las distribuciones de las poblaciones de variables de las cuales se seleccionaron las muestras tienen forma normal.

Pues bien, aquí tenemos esta bonita curva teórica (ya que ninguna curva real es tan perfecta), pero ¿cómo puede ayudarnos a entender lo que significan 105 puntajes individuales?

Comencemos con el papel que la media y la desviación estándar desempeñan en la definición de las características de la curva normal; luego pasaremos al concepto de puntajes estándar.

La media y la desviación estándar

Para comenzar, las curvas pueden diferir marcadamente en su apariencia. Por ejemplo, podemos ver que las dos curvas de la figura 7.5 difieren en su puntaje medio pero no en su variabilidad. En cambio, las dos curvas de la figura 7.6 difieren en su variabilidad pero tienen la misma media.

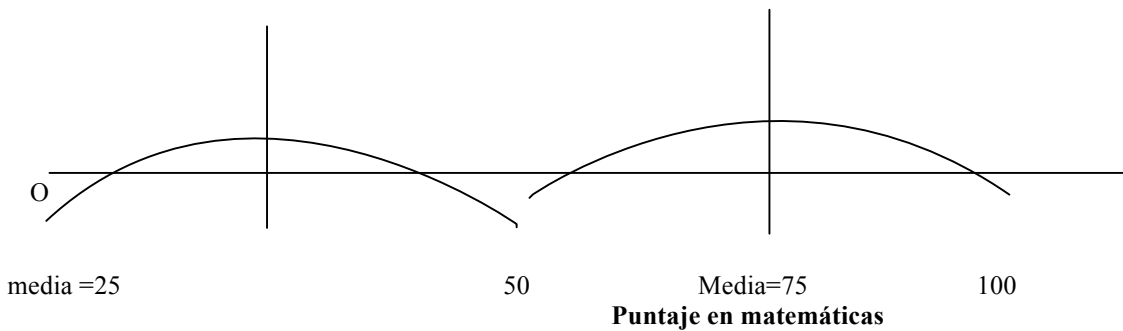


Figura 7.5 Dos distribuciones de puntales que difieren en el puntaje medio pero no en su variabilidad.

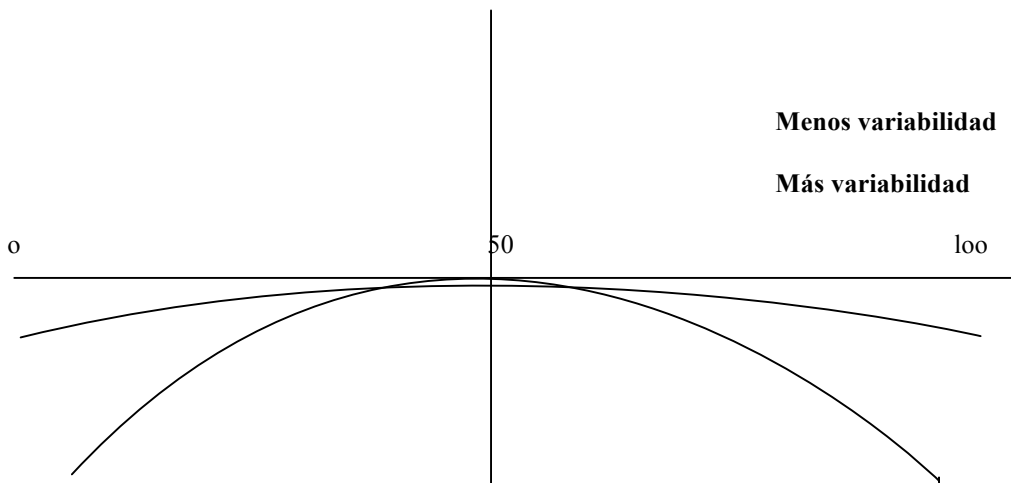


Figura 7.6 Dos distribuciones de puntales que difieren en su variabilidad pero no en su media.

Sea cual sea la forma de la curva o la posición de la media a lo largo del eje X, hay otras condiciones (además de las tres que mencionamos antes) que se cumplen para todas las distribuciones normales. Dichas condiciones son:

La distancia entre la media de la distribución y una unidad de desviación estándar a la izquierda o a la derecha de la media (sin importar qué valor tenga la desviación estándar) siempre abarca aproximadamente 34% (en realidad 34.12%) del área bajo la curva normal, como se muestra en la figura 7.7. Si la media de los puntajes en matemáticas para los 200 estudiantes es de 47.37 y la desviación estándar es de 10.02, 34% de todos los puntajes de la distribución caerán entre los valores de 47.37 y 57.39 (la media más una desviación estándar) y otro 34% caerá entre los valores de 37.35 (la media menos una desviación estándar) y 47.37.

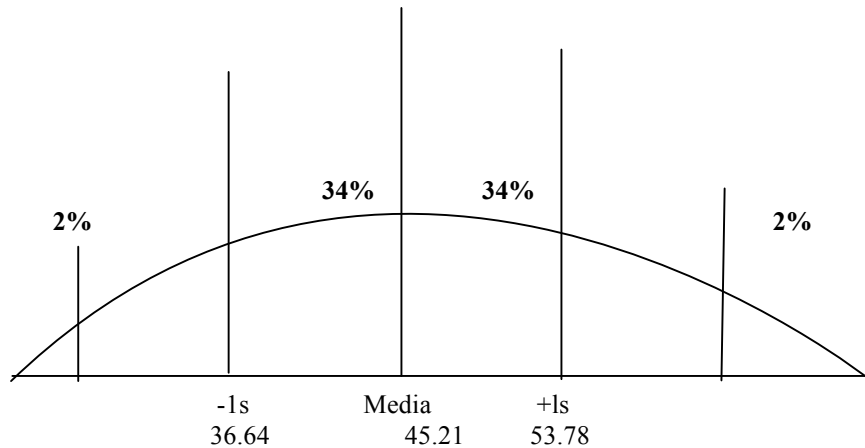


Figura 7.7 La curva normal, indicando los porcentajes y las áreas halo la curva.

Esto es muy interesante si consideramos que el número de 34% es *independiente* del valor real de la media y de la desviación estándar. Este 34% es 34% a causa de la forma de la curva, no del valor de cualquiera de sus miembros o medidas de tendencia central o variabilidad. Si dibujamos una curva normal en un trozo de cartón y recortamos el área entre la media y +1 0-1 desviación estándar y la pesamos, pesará exactamente 34% de lo que pesa el trozo de cartón correspondiente al área bajo toda la curva.

Es evidente que la curva es simétrica. Entonces, en una distribución normal, 68% de todos los puntajes cae entre los valores representados por 1 desviación estándar bajo la media y 1 desviación estándar arriba de la media.

En nuestro ejemplo, esto significa que 68% de los puntajes cae entre los valores de 37.35 y 57.39 (como se observa en la figura 7.7). ¿Qué hay del otro 32%? Buena pregunta. Esos puntajes quedan arriba (a la derecha) de 1 desviación estándar arriba de la media, y abajo (a la derecha) de 1 desviación estándar abajo de la media, en la parte sombreada de la curva de la figura 7.7. En términos más precisos, podemos ver que diferentes porcentajes de los puntajes caen dentro de diferentes fronteras. Puesto que la curva tiene una pendiente, y la cantidad de área disminuye a medida que nos alejamos de la media, no debe sorprendernos de que la probabilidad de que un puntaje caiga cerca de los extremos de la distribución sea menor que la probabilidad de que caiga cerca de la parte media.

Cálculo y uso de los puntajes z

Hemos visto en varios puntos de este capítulo cómo las distribuciones difieren unas de otras primordialmente en función de los valores de la media y de la desviación estándar. Para sacar provecho de la información que se obtiene de las diferentes distribuciones, se necesita un método que tome en cuenta dichas diferencias. Bienvenido a los puntajes estándar.

Los **puntajes estándar** son puntajes que tienen el mismo punto de referencia y la misma desviación estándar. El tipo de puntajes estándar que vemos con mayor frecuencia una bibliografía se denomina **puntaje z** y es el resultado de dividir la desviación de un puntaje individual respecto a la media, entre la desviación estándar, utilizando la fórmula siguiente:

$$z = \frac{(X-x)}{s}$$

- donde z = puntaje z
 X = puntaje bruto
 x = media
 s = desviación estándar

Por ejemplo, si la desviación estándar es 10, el puntaje bruto es 110 y la media de la distribución de puntajes es 100, el puntaje z es

$$z = \frac{(110- 100)}{10} = +1.0$$

En la **tabla 7.5** se muestran los puntajes brutos originales y los puntajes z para el conjunto de 10 puntajes en la prueba de matemáticas que presentamos en la tabla 7.4. Cualquiera

Puntaje original	X-x	Puntal. z [X - x/s)
13	-.4	-.2
14	.6	.31
15	1.6	.82
12	-1.4	-.71
13	-.4	-.2
14	.6	.31
13	-.4	-.2
16	2.6	1.33
15	1.6	.82
9	-4.4	-2.24
Media = 13.4		
s=1.96		

Tabla 7.5 Cálculo de puntajes z a partir de las puntuales brutas.

puntaje bruto que sea mayor que la media tendrá un puntaje z positivo, y cualquier puntaje bruto menor que la media tendrá un puntaje z negativo. Por ejemplo, un puntaje bruto de 13 tiene el puntaje z equivalente de -0.2, que estaría un poco por debajo de la media de 13.4. Un puntaje igual a la media tiene un puntaje z de cero. Un puntaje que está una desviación estándar arriba de la media (15.36) tiene un puntaje z de 1, etcétera.

La aplicación más útil de estos puntajes estándar es para comparar los puntajes de distribuciones que son diferentes entre sí. En la tabla 7.6 se muestra un ejemplo sencillo. El puntaje promedio en matemáticas en el grupo de Sara fue de 90 y la desviación estándar fue de 2. Ella recibió un puntaje de 92, que corresponde a un puntaje z de 1. En la clase de Miguel, el puntaje promedio fue el mismo y él recibió el mismo puntaje absoluto que Sara, pero la desviación estándar fue dos veces mayor (4), lo que hace que su puntaje z $(92-90/4)$ sea 0.5. Podemos ver que, aunque ambos alumnos recibieron el mismo puntaje bruto, el puntaje de Sara está "más arriba" del puntaje de Miguel, lo que indica que su desempeño fue mejor si se utiliza el mismo estándar. ¿Por qué fue mejor su desempeño? Relativo al resto de los miembros de su grupo, ella obtuvo un puntaje más alto. La dispersión en su grupo fue menor; lo que indica que el mismo puntaje absoluto (que ambos alumnos recibieron) los situó en un lugar distinto dentro de cada distribución.

Es por esta razón que los puntajes brutos no deben simplemente sumarse y compararse sus promedios. En vez de ello, se deben utilizar los puntajes z (o algún otro tipo de puntaje estándar) como base de comparación cuando se están considerando puntajes de diferentes distribuciones.

	Media del grupo	s del grupo	Puntaje en la prueba	Puntaje Z
Sara	90	2	92	1
Miguel	90	4	92	.5

Tabla 7.6 Comparación de puntajes z y puntajes brutos.

Lo que realmente significan los puntajes z

Ya sabemos que un puntaje z representa una posición específica a lo largo del eje X. Por ejemplo, en una distribución con una media de 100 y una desviación estándar de 10, un puntaje z de 1 representa el puntaje bruto 110. Asimismo, diferentes puntajes z corresponden a diferentes posiciones a lo largo del eje X de la curva normal. Puesto que ya conocemos el porcentaje del área que cae entre ciertos puntos a lo largo del eje X (véase la figura 7.7), se cumplen afirmaciones como las siguientes:

- 84% de los puntajes quedan por debajo de un puntaje z de +1.0 (50% que queda debajo de la media más 34% que queda entre la media y el puntaje z 1).
- 16% de todos los puntajes quedan por arriba de un puntaje z de +1.0 (puesto que el área total debe ser igual a 100%, y 84% queda

debajo de un puntaje z de $+1.0$).

Podemos hacer este tipo de afirmaciones acerca de la relación entre los puntajes z en cualquier punto a lo largo de la distribución, dado un conocimiento del área correspondiente que se incorpora entre puntos a lo largo del eje. Por ejemplo, utilizando tablas especiales, podríamos determinar que en cualquier distribución normal de 50 puntajes, el número de puntajes que caen entre 1.5 y 2.5 desviaciones estándar por arriba de la media es de unos 3 puntajes, o sea, 6% del total.

Estos números se expresan como porcentajes, que se pueden considerar también una expresión de probabilidad. En otras palabras, la probabilidad de que alguien obtenga un puntaje situado entre 1.5 y 2.5 desviaciones estándar es de .06, o 6 de 100. Asimismo, la probabilidad de que alguien obtenga un puntaje mayor que la media es de .50, o sea, 50%. En el siguiente capítulo estudiaremos esta idea de asignar un valor (en este caso un porcentaje o una probabilidad) a un resultado (en este caso un puntaje) como parte del papel de la inferencia en el proceso de investigación.

Algunas personas en verdad gustan de recopilar datos, mientras que otras consideran esta labor tediosa y cansada. En algo todos están de acuerdo: es un trabajo duro. todo el trabajo rinde frutos cuando comenzamos a armar los datos para formar un cuerpo de información que tiene sentido. Recopilamos los datos, luego los organizamos, y después les aplicamos algunas de las herramientas descriptivas fundamentales de las que hablamos en este capítulo para comenzar a entenderlos. Falta mucho para terminar, pero al menos usted ya tiene una idea de hacia dónde van sus datos.

Ejercicios

1. Usted está encargado de un proyecto que investiga los efectos de las diferencias de género (masculino *versus* femenino) sobre los puntajes en la prueba de lectura de alumnos de primer, tercer y quinto año en tres diferentes distritos escolares. Diseñe un formato de recopilación de datos que tenga en cuenta las variables independientes y dependientes.

- género
- distrito escolar

No olvide dejar espacio en la forma para información importante como las iniciales de la persona que llenó la forma, la fecha de recopilación de los datos, el número de identificación del participante y cualesquier otros comentarios que sean necesarios.

2. La media de una muestra de 10 puntajes es 100 y la desviación estándar es 5. Para los puntajes brutos siguientes, calcule el puntaje z .

- a. 101
- b. 112
- c. 97

Para los puntajes z siguientes, invierta el cálculo y calcule el puntaje bruto correspondiente.

- a. -0.5
- b. 1.1
- c. 2.12

¿Por qué podríamos preferir trabajar con puntajes z en lugar de puntajes brutos? ¿Cuál es el propósito primordial de los puntajes estándar?

3. Para el conjunto siguiente de 10 puntajes, calcule el intervalo, la desviación estándar y la varianza.
5, 7, 3, 4, 5, 6, 7, 2, 5, 3

4. Clara y Noé son excelentes estudiantes. Los resultados de sus pruebas de matemáticas y ciencias fueron:

	Prueba de matemáticas	Prueba de ciencias
Puntaje de Clara	87	92
Puntaje de Noé	78	95
Media del grupo	68	84
Desviación estándar del grupo	8.5	11.5

- a. Calcule los puntajes estándar (puntajes z) para Clara y Noé en matemáticas.
- b. Si un puntaje z mayor implica un "mejor" puntaje, ¿quién recibió la calificación más alta y en qué prueba?
- c. ¿Quién de los dos es mejor estudiante globalmente?

5. Si alguien recibe un puntaje z de 0 ¿qué tan bueno fue su desempeño en comparación con los demás estudiantes del grupo?

6. Calcule la media y la mediana para el siguiente grupo de puntajes:

1, 2, 3, 4, 10

a. En este caso, ¿por qué la mediana es un promedio más útil que la media? b. ¿Por qué no usaría usted la moda como un promedio?

7. Cuando se informa en los medios de comunicación del ingreso promedio de 105 ciudadanos de un país, ¿qué cree usted que se esté citando: la media, la mediana o la moda?
8. Lance un par de dados y registre los resultados después de cada uno de 10 lanzamientos. ¿Algunos totales ocurren con mayor frecuencia que otros? ¿Cree usted que su distribución de puntajes es representativa? ¿Si grafica los resultados, tendrá una curva con forma parecida a la de una curva normal?
9. Una prueba de IQ tiene una media de 100 y una desviación estándar de 15. ¿Qué porcentaje de los niños tendrían un IQ de 1150 más según esta prueba?
10. Mencione los tres tipos de medidas de tendencia central. Defina cada medida.
11. Determine la media, la mediana y la moda para los grupos siguientes:

Grupol	Grupo2
1	3
1	2
1	10
4	3
3	7
5	5

CAPITULO OCHO

Introducción a la estadística inferencial

En Neil J. Salkind “Métodos de Investigación”, Prentice Hall, México, 1998.

Lo que aprenderá en este capítulo

- Por qué el proceso inferencial es importante en investigación
- El papel del azar en cualquier empresa científica
- Qué es la significancia estadística, y por qué es importante
- Qué son los errores tipo I y tipo II, y qué importancia tienen
- Los pasos para llevar a cabo una prueba de significancia estadística
- Algunos tipos básicos de pruebas estadísticas, y cómo se usan
- Un poco acerca de la estadística multivariada y su aplicación
- Un poco más acerca del análisis factorial y su aplicación
- La diferencia entre significancia y significado
- El empleo del metanálisis en las investigaciones sobre la conducta y las ciencias sociales

El entendimiento de las medidas de tendencia central, la variabilidad y el funcionamiento de la curva normal proporciona las herramientas necesarias para describir las características de una muestra. Estas herramientas son además un fundamento excelente para tomar decisiones informadas acerca de la exactitud con que los datos recopilados reflejan la validez de la hipótesis que se está probando.

Una vez que se ha descrito una muestra de datos, como se explicó en el capítulo 7, el siguiente paso es aprender a utilizar esta información descriptiva para hacer una inferencia desde la muestra pequeña de la cual se recopilaron los datos, a la población grande de la cual se escogió originalmente la muestra.

Estadística inferencial

Mientras que la estadística descriptiva sirve para describir las características de una muestra, la estadística inferencial sirve para inferir algo acerca de la población de la cual se extrajo la muestra, con base en las características de esta última. En varios puntos a lo largo de la primera mitad del libro he hecho hincapié en que una de las características distintivas de la buena investigación científica es la selección de una muestra de tal manera que sea representativa de la población de la cual se seleccionó. Cuanto más representativa sea la muestra, más podremos confiar en los resultados basados en la información obtenida de la muestra. La noción de inferencia se basa en el supuesto

de que podemos seleccionar una muestra de modo tal que se maximice dicha representatividad. Entonces, el proceso se convertirá en un proceso inferencial, en el que inferimos de la muestra pequeña a la población grande con base en los resultados de pruebas (y experimentos) realizados con la muestra.

Como funciona la inferencia

Sigamos los pasos de un proyecto de investigación para ver cómo podría funcionar el proceso inferencial.

Por ejemplo, un investigador desea averiguar si existe una diferencia significativa entre los adolescentes de sexo masculino y femenino en cuanto a la forma como resuelven dilemas morales. Haciendo un repaso de los pasos generales del proceso de investigación que vimos en el capítulo 1, he aquí una posible secuencia de acciones.

1. El investigador escoge muestras representativas tanto de hombres como de mujeres de tal manera que las muestras representan las poblaciones de las que se extrajeron.
2. A cada adolescente se le administra una prueba para evaluar su nivel de desarrollo moral.
3. El puntaje medio del grupo de hombres se compara con el puntaje medio del grupo de mujeres utilizando alguna prueba estadística.
4. Se llega a una conclusión en cuanto a si la diferencia entre los puntajes es el resultado del azar (es decir, que algún factor distinto del género es la causa de la diferencia) o es resultado de diferencias "verdaderas" entre los dos grupos.
5. Se llega a una conclusión acerca del papel que el género desempeña en el desarrollo moral de la población de la cual se extrajo originalmente la muestra. En otras palabras, se hace una inferencia acerca de la población con base en los resultados de un análisis de los datos de la muestra.

El papel del azar

De hecho, si no se sabe algo más acerca de la relación entre las variables en cuestión, el azar siempre es la explicación más atractiva de cualquier relación que pudiera existir. Por ejemplo, antes de eliminar todas las posibles causas de cualesquier diferencias entre el desarrollo moral de los dos grupos de adolescentes, la explicación más atractiva es que, si los dos grupos difieren, es por casualidad. ¿Qué es el azar? Es la ocurrencia de variabilidad que no se puede justificar por cualquiera de las variables que se están estudiando. Es por ello que no podemos partir del supuesto de que cualquier diferencia que observemos entre hombres y mujeres se debe a las diferencias de género. No hay indicios que apoyen tal supuesto.

Nuestro papel principal como científicos es *minimizar el papel que el azar podría desempeñar en el entendimiento de la relación entre variables*. Esto se hace primordialmente controlando las diferentes fuentes de variabilidad (diferencias como experiencia previa, edad, etc.) que pudieran existir.

Explicaremos cómo controlar las diversas fuentes de error (o las explicaciones alternativas de los resultados) en el capítulo 10. Por ahora, expliquemos por qué podemos examinar una muestra relativamente pequeña de observaciones y hacer una inferencia respecto a una población mucho mayor. La técnica (y la justificación subyacente) es en verdad fascinante y sustenta gran parte de los informes cotidianos de resultados científicos. ¡Nada más piense en la precisión que los encuestadores profesionales deben aplicar a su selección de una muestra de unos 1 200 adultos que representan a más de 160 millones de adultos!

El teorema del límite central

El eslabón absolutamente crucial entre los resultados que obtenemos de la muestra y la capacidad para generalizar esos resultados a la población es el supuesto de que un muestreo repetido de la población producirá un conjunto de puntajes representativo de la población. Si esto no sucede, no podremos aplicar muchas (si es que no todas) las pruebas de la estadística inferencial.

Recuerde la pregunta que hicimos antes: "¿Cómo sabemos si la distribución de la población de la cual escogimos una muestra es normal?" La respuesta es que no lo sabemos, porque nunca podremos examinar o evaluar realmente las características de toda la población. Lo que es más, en cierto sentido ni siquiera nos debe importar (¡qué horror!), en virtud del **teorema del límite central** que dice que, *sea cual sea la forma de la distribución de la población (sea normal o no), ¡las medias de todas las muestras seleccionadas de la población tendrán una distribución normal!* Esto implica que, incluso si la distribución de una población de puntajes tiene forma de U invertida (exactamente lo opuesto a una curva de campana), si escogemos varias muestras de 300 más observaciones cada una de esa población, las medias de esas muestras tendrán una distribución normal. Comprobaremos esto en un momento más, pero primero recline su silla y medite un poco sobre lo que esta observación realmente significa para la aplicación de estos principios al mundo real de la investigación, donde la verdadera forma de la distribución de los puntajes de una población no es normal, ni tiene forma de campana.

Lo más importante es que esta observación implica que no necesitamos conocer la distribución de puntajes dentro de la población para generalizar los resultados de la muestra a la población. Esto es muy importante, pero también es evidente que si no sucediera así sería muy difícil, si no imposible, inferir de una muestra a la población de la cual se extrajo.

Una de las claves para tener éxito al aplicar este teorema es que el tamaño de muestra sea mayor que 30. Si el tamaño es menor, podría ser necesario aplicar estadística no paramétrica o independiente de la distribución (la que aplica a "curvas no normales").

Un ejemplo del teorema del límite central

En la tabla 8.1 se muestra una población de 100 valores que van desde 1 hasta 5, y la figura 8.1 es una gráfica de su distribución (puntaje *versus* frecuencia). La media de toda la

1	5	5	2	5
2	3	2	5	1
1	5	2	1	1
4	4	5	1	1
1	5	5	5	1
1	3	5	1	1
4	5	4	1	4
1	3	1	4	5
5	2	5	5	5
1	1	5	3	5
2	5	2	2	1
5	2	5	5	1
5	5	5	5	1
2	4	2	2	1
1	1	1	1	2
4	4	3	4	5
1	1	4	5	1
5	4	1	5	4

Tabla 8.1 Una población de 100 puntajes, que es una minipoblación. Puede verse qué aspecto tiene la distribución de estos puntajes si se grafican éstos (que varían entre 1 y 5) contra el número de veces que aparece cada puntaje.

población es de 3.0. Como puede verse, la distribución tiene forma de U. En el mundo real nunca podría observarse toda la población (si se pudiera, ¿para qué nos serviría la inferencia?), pero para fines ilustrativos tengamos un poco de fe y supongamos que es posible.

Se selecciona una muestra de 10 valores de esta población (aleatoriamente) y se calcula la media. Su valor (Media 1) es 4. Ahora se selecciona otra muestra (Media 2) y así sucesivamente, hasta haber calculado las medias de 30 muestras distintas de tamaño 10. Estas 30 medias (redondeadas) se listan en la tabla 8.2. Si graficamos las medias (como si fueran una distribución de puntajes), vemos que su distribución se acerca a la normalidad, como se aprecia en la figura 8.2. Así, a partir de una población cuyos puntajes tenían una distribución opuesta (es decir, con forma de U, figura 8.1), a una curva normal, podemos generar una distribución normal de valores. Además, la "media de todas las medias" (el promedio de las 30 medias de muestra distintas) es muy cercana (2.76) a la media de la población original (¿recuerda que era 3.00?) de la cual se sacaron. ¿Coincidencia? Para nada. ¿Sorprendente? En absoluto. Es simplemente la potencia del teorema del límite central.

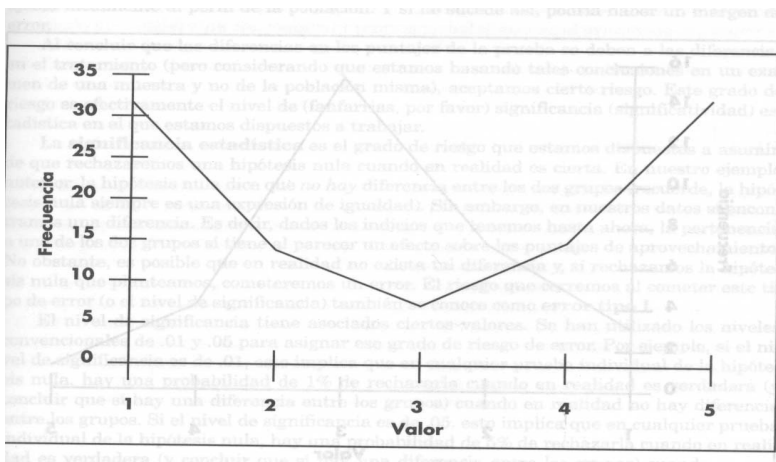


Figura 8.1 grafica de 100 datos puntuales

Tabla 8.2 las medias, cada una generada de una muestra de tamaño 10

Este teorema es importante: ilustra lo potente que puede ser la estadística inferencial porque permite tomar decisiones con base en las características de una curva normal aun si la población de la cual se extrajo la muestra no *es normal*. Este hecho por sí solo ofrece una flexibilidad enorme y en muchos sentidos es la piedra angular del método experimental. Sin la capacidad para inferir, habría que aplicar la prueba a toda la población, tarea poco razonable y poco práctica.

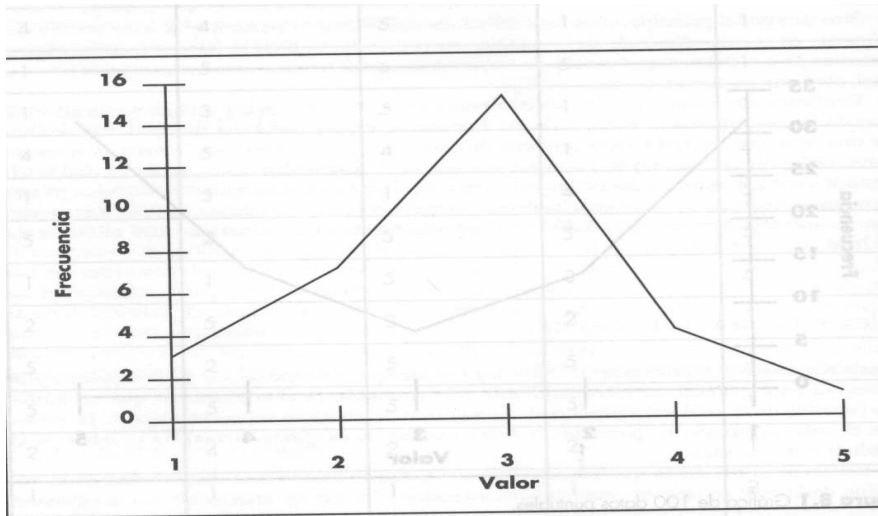


Figura 8.2 Una distribución de medias.

Figura 8.2 Una distribución de medias.

La idea de la significancia estadística

Dado que los muestreos son imperfectos (en cuanto a que nunca podremos seleccionar una muestra de sujetos cuyo perfil coincida exactamente con el de la población), introducimos cierto error (error de muestreo) en el proceso de muestreo. Además, dado que las hipótesis no se pueden aplicar la prueba directamente con las poblaciones (sencillamente porque no es práctico; las poblaciones son demasiado grandes), es posible que las inferencias tampoco sean perfectas. También, cabe la posibilidad de que las inferencias sean erróneas al concluir que dos grupos son diferentes unos de otros (lo cual podríamos concluir por los datos de la muestra) cuando en realidad (que es la condición que realmente prevalece en la población) no lo son.

Por ejemplo, digamos que a un investigador le interesa averiguar si existe una diferencia en el aprovechamiento académico de niños que participaron en un programa preescolar y el de aquellos que no lo hicieron. La hipótesis nula es que los dos grupos son iguales en cuanto a alguna medida del aprovechamiento.

La hipótesis de investigación es que el puntaje medio del grupo de niños que participaron en el programa es más alto que el puntaje medio del grupo de niños que no participaron en el programa.

Como buenos investigadores, nuestra tarea es demostrar (lo mejor que podamos) que cualesquier diferencias que existan entre los dos grupos se deben *únicamente* a los efectos de la experiencia preescolar y a ningún otro factor o combinación de factores. Mediante diversas técnicas que describiremos en el capítulo 10, controlamos o eliminamos todas las posibles fuentes de diferencias, como la influencia de la educación de los padres, el número de niños en la familia, etc. Una vez eliminadas éstas y otras posibles variables explicativas, la única explicación alternativa que queda para las diferencias es el efecto de la experiencia preescolar. Pero, ¿podemos estar absolutamente seguros de ello? No, no podemos. ¿Por qué? Porque no estamos seguros de estar aplicando pruebas a una muestra que se ajusta idealmente al perfil de la población. Y si no sucede así, podría haber un margen de error.

Al concluir que las diferencias en los puntajes de la prueba se deben a las diferencias en el tratamiento (pero considerando que estamos basando tales conclusiones en un examen de una muestra y no de la población misma), aceptamos cierto riesgo. Este grado de riesgo es

efectivamente el nivel de (fanfarrias, por favor) significancia (significatividad) estadística en el que estamos dispuestos a trabajar.

La significancia estadística es el grado de riesgo que estamos dispuestos a asumir de que rechazaremos una hipótesis nula cuando en realidad es cierta. En nuestro ejemplo anterior la hipótesis nula dice que no *hay* diferencia entre los dos grupos (recuerde, la hipótesis nula siempre es una expresión de igualdad). Sin embargo, en nuestros datos sí encontramos una diferencia. Es decir, dados los indicios que tenemos hasta ahora, la pertenencia a uno de los dos grupos sí tiene al parecer un efecto sobre los puntajes de aprovechamiento. No obstante, es posible que en realidad no exista tal diferencia y, si rechazamos la hipótesis nula que planteamos, cometeremos un error. El riesgo que corremos al cometer este tipo de error (o el nivel de significancia) también se conoce como error **tipo 1**.

El nivel de significancia tiene asociados ciertos valores. Se han utilizado los niveles convencionales de .01 y .05 para asignar ese grado de riesgo de error. Por ejemplo, si el nivel de significancia es de .01, esto implica que en cualquier prueba individual de la hipótesis nula, hay una probabilidad de 1% de rechazarla cuando en realidad es verdadera (y concluir que sí hay una diferencia entre los grupos) cuando en realidad no hay diferencia entre los grupos. Si el nivel de significancia es de .05, esto implica que en cualquier prueba individual de la hipótesis nula, hay una probabilidad de 5% de rechazarla cuando en realidad es verdadera (y concluir que sí hay una diferencia entre los grupos) cuando en realidad no hay diferencia entre los grupos. Observe que el nivel de significancia está asociado a una prueba independiente de la hipótesis nula, y no es correcto decir que "En 100 pruebas de la hipótesis nula, sólo cometeré un error o cinco

En un informe de investigación, la significancia estadística suele representarse como $p < .05$ (léase: "la probabilidad de observar ese resultado es menor que .05"), y a menudo se expresa en un informe o artículo simplemente como "significativo en el nivel .05".

Existe otro tipo de error que podemos cometer mismo que se muestra en la tabla 8.3 junto con el error tipo 1. Cometemos un error **tipo II** cuando sin querer aceptamos una hipótesis nula falsa. Por ejemplo, si en realidad hay diferencias entre las poblaciones representadas por los grupos de muestra, pero erróneamente concluimos que no hay diferencias.

Si usted	Cuando es	Usted está
rechaza la hipótesis nula	verdadera	cometiendo un error tipo 1
rechaza la hipótesis nula	falsa	en la correcto
rechaza la hipótesis nula	verdadera	cometiendo un error tipa II
rechaza la hipótesis nula	falsa	en la correcta

Tabla 8.3 Los tipos de errores que pueden cometerse al trabajar can una hipótesis nula.

Idealmente, queremos minimizar ambos tipos de errores, 1 y II, pero no siempre es fácil, y podría estar filera de nuestro control. Tenemos control absoluto sobre el nivel de error tipo 1, el riesgo que estamos dispuestos a correr (porque somos nosotros quienes fijamos el nivel de significancia). Los errores tipo II no se controlan de manera tan directa, pues están relacionados con factores tales como el tamaño de la muestra. Los errores tipo II son muy sensibles al número de sujetos en una muestra, y a medida que ese número aumenta el error tipo II disminuye. En otras palabras, a medida que las características de la muestra se acercan más a las de la población (lo que se logra aumentando el tamaño de la muestra), disminuye la probabilidad de aceptar una hipótesis nula falsa.

Pruebas de significancia

Lo que mejor hace la estadística inferencial es ayudarnos a tomar decisiones acerca de las poblaciones con base en la información que poseemos de las muestras. Una de las herramientas más útiles para hacer esto es una prueba de significancia estadística que se puede aplicar a diferentes tipos de situaciones, dependiendo de la naturaleza de la pregunta que se está planteando y de la forma de la hipótesis nula.

Por ejemplo, ¿quiere usted examinar la diferencia entre dos grupos, digamos silos niños obtienen puntajes significativamente distintos de los de las niñas en alguna prueba? ¿O desea examinar la relación entre dos variables, como el número de niños de una familia y el puntaje promedio en pruebas de inteligencia? Los dos casos requieren diferentes estrategias, pero ambos se reducirán a una prueba de la hipótesis nula empleando una prueba de significancia estadística específica.

Cómo funcionan las pruebas de significancia

Las pruebas de significancia se basan en el hecho de que cada tipo de hipótesis nula (como $H_0 : \mu_{1980} = \mu_{1984}$) que representa la ausencia de diferencia entre las medias de dos muestras) tiene asociado un tipo de estadística específico. La estadística tiene asociada una distribución de valores que sirve para comparar lo que los datos de muestra revelan y lo que cabría esperar del azar. Una vez más, el azar es la más creíble de todas las explicaciones si no tenemos indicios que indiquen otra cosa.

He aquí los pasos generales a seguir en la aplicación de una prueba estadística a cualquier hipótesis nula:

1. *Expresión de la hipótesis nula.* ¿Recuerda que la hipótesis nula es una expresión de igualdad? La hipótesis nula es la situación "verdadera" si no tenemos otra información que nos permita emitir un juicio. Por ejemplo, si nada sabemos acerca de la relación entre la memoria a largo plazo y la práctica diaria, supondremos que no hay relación entre ellas. Tal vez eso no sea lo que desea probar como

hipótesis, pero siempre es el punto de partida.

2. *Establecer el nivel de riesgo (o el nivel de significancia o error tipo 1) asociado a la hipótesis nula.* Toda hipótesis de investigación conlleva cierto grado de riesgo de cometer un error tipo 1. Cuanto menor sea este error (digamos .01 en lugar de .05), menos riesgo estaremos dispuestos a correr. Ninguna prueba de hipótesis está totalmente exenta de riesgo porque nunca conocemos realmente la relación "verdadera" entre variables.

3. *Selección de la prueba estadística apropiada.* Cada hipótesis nula tiene asociada una prueba estadística específica. Usted puede aprender cuál prueba está relacionada con cuál tipo de pregunta con mayor detalle en los cursos de estadística que se imparten en su escuela.

4. *Calculo del valor de la prueba estadística (llamado valor obtenido).* El valor obtenido es el resultado de una prueba estadística específica. Por ejemplo, existen pruebas estadísticas para la significancia de la diferencia entre los promedios de dos grupos, para la significancia de la diferencia respecto a cero de un coeficiente de correlación, y para la significancia de la diferencia entre dos proporciones.

5. *Determinación del valor requerido para rechazar la hipótesis nula empleando la tabla apropiada de valores críticos para la estadística en cuestión.* Cada prueba estadística junto con el tamaño de grupo y el riesgo que estamos dispuestos a correr) tiene asociado un valor crítico. Éste es el valor mínimo que cabría esperar tuviera la prueba estadística si la hipótesis nula es realmente verdad.

6. *Comparación del valor obtenido con el valor crítico.* Éste es el paso crucial. Aquí el valor que obtuvimos de la prueba estadística (el que calculamos) se compara con el valor que esperaríamos obtener si el azar fuera el único factor (el valor crítico).

7. *Si el valor obtenido es mas extremo que el valor critico, no es posible aceptar la hipótesis nula.* Es decir; la expresión de igualdad de la hipótesis nula (que refleja el azar) no es la explicación más atractiva de las diferencias detectadas. Aquí es donde brilla realmente el método inferencial. Sólo si el valor obtenido es más extremo que el azar (lo que implica que el valor de la prueba estadística no es resultado de alguna fluctuación fortuita) podemos decir que cualesquier diferencias observadas no se deben al azar y que la igualdad expresada por la hipótesis nula no es la explicación más atractiva de las diferencias halladas. Más bien, las diferencias se deben al tratamiento.

8. *Si el valor obtenido no excede el valor critico, la hipótesis nula es la explicación mas atractiva.* Si no podemos demostrar que la diferencia que observamos se debe a otra cosa que al azar (digamos, el tratamiento), entonces la diferencia seguramente se debe al azar o a algo sobre lo que no se tiene control. En otras palabras, la hipótesis nula es la mejor explicación.

Sigamos estos pasos en el contexto de un ejemplo de aplicación de una prueba de significancia.

Prueba t para medias independientes

La prueba t para medias independientes es una prueba inferencial de la significancia de la diferencia entre dos medias basadas en dos grupos independientes, sin relación una con otra. Se trata de dos grupos distintos, como hombres y mujeres, o quienes recibieron un tratamiento y quienes no lo recibieron.

Chuangsheng Chenn y Harold Stevenson (1989) examinaron las diferencias culturales entre 3 500 niños de escuela primaria y sus padres y maestros en Beijing (China), Chicago, Minneapolis, Sendai (Japón) y Taipei. Una de las hipótesis de investigación asociadas a este gran conjunto de estudios fue que la cantidad de tarea escolar que los niños hacían en casa (según la estimación de las madres de los niños) cambió a lo largo del periodo de estudio de cuatro años de 1980 a 1984.

Los ocho pasos que acabamos de describir (utilizando este estudio como ejemplo) son los siguientes:

1. *Expresión de la hipótesis nula.* En este caso la hipótesis nula es que no hay diferencia entre la cantidad promedio de tiempo dedicada a hacer la tarea en 1980 y el tiempo dedicado a hacer la tarea en 1984. Utilizando símbolos, la hipótesis nula se expresa así:

$$H_0 : \mu_{1980} = \mu_{1984}$$

donde H_0 representa la hipótesis nula

μ_{1980} = el promedio de la población para los niveles de tarea de 1980

μ_{1984} = el promedio de la población para los niveles de tarea de 1984

Recuerde que puesto que las hipótesis nulas siempre se refieren a poblaciones, se utilizan parámetros como μ para representar la media, en lugar de X

2. *Establecer el nivel de riesgo (o el nivel de significancia o error tipo 1) asociado a la hipótesis nula.* Por convención, se asigna un valor de .05 o .01. En este caso usaremos el valor .05.

3. *Selección de la prueba estadística apropiada.* La prueba estadística apropiada para esta hipótesis nula es la prueba t entre medias independientes. Las medias son independientes porque son promedios calculados a partir de grupos distintos.

4. *Calculo del valor de la prueba estadística (valor obtenido).* En el presente estudio, el valor de la prueba estadística para la comparación de 320 estimaciones, hechas por las madres, del tiempo dedicado a hacer la tarea en 1980 y 1984 fue 2.00. Este fue el resultado de aplicar la fórmula mencionada en el Paso 3. Tomamos este valor directamente del artículo publicado.

Determinación del valor (llamado valor critico) requerido para rechazar la hipótesis nula empleando la tabla apropiada de valores críticos para la estadística en cuestión. Para determinar el valor crítico, es preciso consultar una tabla para la estadística empleada. La tabla 8.4 es

una porción de semejante tabla.

Grados de libertad	Valor crítico para rechazar la hipótesis nula en el nivel de significancia de .05	Valor crítico para rechazar la hipótesis nula en.' niv.' de significancia de .01
40	2.021	2.704
60	2.000	2.660
120	1.960	2.617

Tabla 8.4 Tabla parcial de valores críticos utilizados para probar un valor t para la diferencia entre los promedios de dos grupos independientes. Observe que al aumentar el tamaño de la muestra, el valor crítica disminuye.

Para determinar el valor crítico que una prueba estadística necesita para alcanzar significancia, hay que saber dos cosas: el nivel de significancia en el que se está probando la hipótesis de investigación (.05 en este caso) y los grados de libertad, un parámetro que refleja el tamaño de la muestra (320 en este caso). Necesitamos conocer el tamaño de la muestra porque el valor crítico cambia al cambiar el tamaño de la muestra. ¿Puede usted imaginarse por qué? Porque a medida que el tamaño de la muestra aumenta, la muestra se parece más a la población, y la diferencia que necesitamos entre el valor obtenido y el crítico para rechazar la hipótesis nula se reduce.

Examine la tabla 8.4 y utilice esta información para determinar el valor crítico. Lea hacia abajo la columna rotulada grados **de libertad** hasta acercarse lo más posible a 320 (que es 120). Ahora lea horizontalmente hacia la columna para el nivel de significancia de .05. Puesto que la hipótesis no señala alguna dirección para la diferencia, se trata de una prueba *de dos colas*, o *no direccional*. En el cruce de 120 grados de libertad y el nivel de .05 vemos que se necesita el valor crítico de 1.960 para rechazar la hipótesis nula.

Éste es el valor t que esperaríamos únicamente por el azar, dado el tamaño de la muestra y el nivel de significancia en el que queremos probar la hipótesis de investigación.

6. *Comparación del valor obtenido con el valor crítico.* Aquí, los dos valores de interés son el valor obtenido ($t = 2.00$) y el valor crítico ($t = 1.96$).

7. *Si el valor obtenido es mas extremo que el valor crítico, no es posible aceptar la hipótesis nula.* Es decir, esta expresión de igualdad (que refleja el azar) no es la explicación más atractiva para las diferencias observadas. En este caso, el valor obtenido es mayor que el valor crítico. En otras palabras, la probabilidad de que este valor t sea el resultado exclusivamente del azar es menor que .05 (o 5 de 100) en cualquier prueba individual de la hipótesis nula. La conclusión basada en los datos de muestra, entonces, es que existe una diferencia en el número promedio de minutos que se dedican a hacer la tarea, entre 1980 y 1984. ¿Cuál es la naturaleza de la diferencia? Un examen de las medias (252 minutos por semana en 1980 *versus* 305 minutos por semana en 1984) indica que los niveles de tarea aumentaron, resultado congruente con la hipótesis de investigación antes mencionada.

8. *Si el valor obtenido no excede el valor crítico, la hipótesis nula es la explicación mas atractiva.* En este caso, el valor obtenido excedió el valor crítico. La hipótesis nula no es la explicación más atractiva ni sostenible de las diferencias.

Entonces, ¿qué significa *realmente* $t(120) = 2.00, p < .05$?

A medida que usted se familiarice con los artículos de revistas científicas y la manera en que están escritos, irá reconociendo declaraciones como la siguiente:

Los resultados fueron significativos en el nivel de .05 ($t(120) = 5.43, p < .01$). Las palabras están claras, pero ¿qué significa cada una de las partes?

- La t representa el tipo de prueba estadística, que en este caso es una prueba t .
- El (120) representa el número de grados de libertad.
- El 5.43 es el valor obtenido, el que se calcula aplicando la prueba t a los resultados del estudio
- La p representa probabilidad.
- El .01 representa el nivel de significancia o la tasa de errores tipo 1.

Una vez que adquiera experiencia en la lectura de estas expresiones gracias a su exposición a estudios terminados y artículos publicados, se le hará muy fácil echar una mirada rápida a los números y reconocer su significado. El formato aquí presentado es hasta cierto punto estándar; el valor de los elementos cambia (como F para una prueba F , 0.05 para un nivel de significancia distinto) pero no su significado.

Diferentes pruebas de significancia

Ya vimos que hay diferentes pruebas de significancia que se pueden aplicar a diferentes tipos de preguntas. En el ejemplo anterior, la prueba de significancia apropiada examinó la diferencia entre los promedios de dos grupos que no tenían relación entre ellos, es decir, eran independientes. Veamos otras pruebas comunes de significancia estadística. Tenga presente que hay más de 100 pruebas distintas que

se pueden aplicar. En la tabla 8.5 presentamos una muestra de algunas de esas pruebas junto con la pregunta de investigación correspondiente, la hipótesis nula y la prueba estadística apropiada.

El propósito de los ejemplos que siguen es que usted conozca algunas de las pruebas de uso más frecuente, con las que probablemente se topará al consultar la bibliografía de investigación. Igualmente, si desea saber más acerca de estas pruebas, considere inscribirse en los cursos de estadística de primer y segundo nivel que seguramente ofrece su escuela.

Examen de las diferencias entre grupos

Acabamos de ver un ejemplo de aplicación de una prueba estadística para examinar la diferencia entre los promedios de dos grupos cuando las mediciones de cada uno de ellos no tienen relación entre sí; es decir, cuando las mediciones son independientes.

La pregunta	La hipótesis nula	La prueba estadística
Diferencias entre grupos		
¿Hay diferencia entre las medias de dos grupos no tratadas?	$H_0: \mu_{\text{grupo1}} = \mu_{\text{grupo2}}$	Prueba t para muestras independientes
¿Hay diferencia entre las medias de dos grupos relacionadas?	$H_0: \mu_{\text{grupo1a}} = \mu_{\text{grupo1b}}$	Prueba t para muestras dependientes
¿Hay una diferencia global entre tres o más medias?	$H_0: \mu_{\text{grupo1}} = \mu_{\text{grupo2}} = \mu_{\text{grupo3}}$	Análisis de varianza (o prueba F)
Relaciones de grupo		
¿Hay alguna relación entre los puntales de dos grupos?	$H_0: r_{xy} = 0$	Prueba t para la significancia del coeficiente de correlación
¿Hay alguna diferencia entre dos coeficientes de correlación?	$H_0: r_{xy} = r_{xz}$	Prueba t para la significancia de la diferencia entre coeficientes de correlación.

Tabla 8.5 Para qué sirven las hipótesis nulas. Para casi cualquier pregunta de investigación que usted desee hacer y que se base en datos de una muestra, existe una hipótesis nula correspondiente.

Otra situación común es aquella en la que los grupos están relacionados unos con otros. Por ejemplo, ¿qué tal si a usted le interesa averiguar qué cambios, si acaso, ocurrieron durante todo el año escolar en los puntajes de competencia en lectura para el mismo grupo de niños? Usted podría administrar la prueba de competencia en septiembre y otra vez en junio. La hipótesis nula es que no hay diferencia entre los puntajes de ambas pruebas. Puesto que los puntajes de las dos pruebas tienen una dependencia mutua, o están correlacionados, y dado que los mismos alumnos están presentando ambas pruebas, la prueba t para medias independientes no es la apropiada. En vez de ella, se debe usar *una prueba t para medias dependientes o correlacionadas*. La diferencia primordial entre estos dos procedimientos es que la prueba para medias dependientes toma en cuenta el grado en que ambos conjuntos de puntajes están relacionados entre ellos.

Por ejemplo, el puntaje medio de 28 muchachos en la prueba de lectura del otoño fue 76.8 con una desviación estándar de 6.5. El puntaje medio para los mismos muchachos en la prueba de lectura de la primavera fue de 82.4 con una desviación estándar de 7.8. ¿Existe una diferencia significativa entre las dos pruebas. Sigamos los pasos antes identificados.

1. Expresión de la hipótesis nula.

$$H_0: \mu_f = \mu_s$$

2. Establecimiento del nivel de riesgo (o el nivel de significancia o el error tipo 1) asociado a la hipótesis nula. Se usará el valor de .01.

3. Selección de la prueba estadística apropiada. La prueba estadística apropiada para esta hipótesis nula es la prueba t entre medias dependientes. Las medias dependen una de otra porque son promedios basados en el desempeño del mismo grupo.

4. Cálculo del valor de la prueba estadística, que es

$$t=2.581$$

5. Determinación del valor necesario para rechazar la hipótesis nula. Utilizando los valores de la tabla 8.6, determinamos el valor crítico tal como hicimos para una prueba de medias independientes. La razón es que se está usando la misma prueba estadística. El número de grados de libertad es $n - 1$ (27) donde n es igual al número de pares de observaciones, que en este caso es 28. Se usa $n - 1$, no n , porque queremos una estimación conservadora del valor de la población, así que subestimamos intencionalmente el tamaño de la muestra (27 en

lugar de 28).

Grados de libertad	Nivel de significancia de una prueba direccional o de una sola cola			
	.05	.025	.01	.005
	Nivel de significancia de una prueba no direccional o de dos colas			
	.10	.05	.02	.01
26	1.706	2.056	2.479	2.779
27	1.703	2.052	2.473	2.771
28	1.701	2.048	2.467	2.763
29	1.699	2.045	2.462	2.756
30	1.697	2.042	2.457	2.750

tabla 8.6 Otra tabla parcial de valores criticos. Puede verse que esta tabla contiene el mismo tipo de información que la tabla 8.4, pero ahora aplicada a pares de observaciones.

Dada esta información y un nivel de significancia de .01, el valor critico necesario para rechazar la hipótesis nula en una prueba de dos colas es de 2.771.

6. Si el valor obtenido es más extremo que el valor critico, no podremos aceptar la hipótesis nula. Es decir, esta expresión de igualdad (que refleja el azar) no es la explicación más atractiva de cualesquier diferencias observadas. En este caso, el valor obtenido de 2.581 no excede el valor critico de 2.771, así que continuamos con el siguiente paso.

7. Si el valor obtenido no excede el valor critico, la hipótesis nula es la explicación más atractiva. La observación basada en los datos de muestra no es lo bastante extrema como para rechazar la hipótesis nula y concluir que si hay una diferencia significativa entre las dos pruebas de lectura. La hipótesis nula de que no hay diferencia entre los dos grupos es la explicación más atractiva. Cualquier diferencia observada (76.8 *versus* 87.4) se debe al error de muestreo.

Examen de las relaciones entre grupos

En el capítulo 9 estudiaremos una prueba de estadística descriptiva llamada coeficiente de correlación (que también mencionamos en el capítulo 5), un índice matemático de la relación entre dos variables. Si nada sabemos acerca de dos variables, llamémoslas X y Y, ¿qué relación esperaríamos que hubiera entre ellas, únicamente por el azar? Puesto que no tenemos razón para creer que estén relacionadas, tenemos que suponer que la relación es 0. Esto es exactamente lo que esperaríamos si el azar fuera el único factor que opera y si las dos variables en cuestión nada tuvieran en común.

La prueba de significancia de una correlación consiste en determinar si el valor del coeficiente, y por ende la relación entre las variables, es significativamente diferente de cero. La hipótesis nula es

$$P_{XY} = 0$$

Por ejemplo, supongamos que usted desea probar la hipótesis de investigación (en el nivel 01) de que la relación entre los puntajes en matemáticas y en lectura (donde el coeficiente de correlación es igual a .13, o $r_{xy} = .13$) es diferente de cero.

El valor de esta prueba estadística es una distribución de puntajes t. Una vez calculado ese valor t, consultamos la misma tabla que utilizamos para las distintas pruebas estadísticas en las que intervienen puntajes t.

Si no podemos rechazar la hipótesis nula, lo que efectivamente estamos diciendo es que no existe relación entre las dos variables. Si no hay una relación significativa (o "real") entre X y Y, entonces ¿cómo puede cualquier coeficiente de correlación (digamos .13) ser diferente de cero? Es sencillo. Se trata del *error de muestreo*. Efectivamente, el valor de .13 no es el verdadero valor que encontraríamos en la población de la cual se extrajo la muestra, sino sólo una función de un muestreo inexacto o impreciso. El error de muestreo es esa amenaza omnipresente, y una de nuestras tareas es separar las diferencias que se deben al error de muestreo, de las diferencias o relaciones verdaderas que existen en la muestra que se está examinando.

Como trabajar con más de una variable dependiente

La pregunta de investigación que usted esta haciendo podría exigirle evaluar más de una variable dependiente. En este caso hay al menos dos técnicas que son mas avanzadas que cualquier otra cosa que hayamos mencionado hasta ahora y le conviene familiarizarse con ellas.

El análisis **de varianza multivariado** (o MANOVA, en inglés, *multivariate analysis of variance*) es una técnica avanzada que examina la ocurrencia de diferencias por grupo en más de una variable dependiente. En muchos sentidos, MANOVA semeja una serie de pruebas t sencillas entre grupos. La principal diferencia entre las dos técnicas es que MANOVA toma en cuenta la relación entre las variables

dependientes. En otras palabras, si las variables dependientes están íntimamente relacionadas entre ellas, sería difícil saber en qué grado una diferencia en la variable dependiente 1 es el resultado de diferencias en la variable dependiente 2. MANOVA separa la contribución única de cada variable dependiente para entender las diferencias entre los grupos, de modo que si existe una diferencia en la variable 1 no se mezcle con cualquier diferencia en la variable 2.

El hecho de que las variables dependientes puedan estar relacionadas entre ellas hace que varias pruebas t por pares sean una amenaza grave para un estudio sólido. Por ejemplo, digamos que usted está probando las diferencias entre el grupo experimental y el de control en cuanto a variables llamadas comprensión, memoria, recuerdo y rapidez de lectura. Como cabría sospechar; todas estas variables están relacionadas unas con otras. Por tanto, una prueba t para evaluar las diferencias en cuanto a rapidez de lectura entre grupos podría indicar que las diferencias son significativas, pero la verdadera razón de la diferencia es que la rapidez de lectura está íntimamente ligada con la comprensión, y es ahí donde radica la verdadera diferencia.

A causa de la interrelación de estas variables, el verdadero error tipo 1 o alfa no es .01 ni .05, sino más bien

$$1-(1-a)^k$$

donde a = tasa de error tipo 1, y
 k = el número de variables.

En este caso, con tres variables, el verdadero error tipo 1 es $1 - (1 - .05)^3 = .14$, que ciertamente es diferente del valor de .05 que habíamos supuesto. En otras palabras, las pruebas t múltiples son riesgosas porque inflan artificialmente el nivel del error tipo 1 que creemos estar manejando. ¿La solución? Utilizar algún tipo de técnica que controle las relaciones entre las variables dependientes. Inicialmente, dicha técnica es MANOVA, seguida de algún tipo de procedimiento *post-hoc* que compare las medias unas con otras y controle el nivel de error tipo 1.

El análisis **factorial** es otra técnica avanzada que permite al investigador reducir el número de variables que representan un constructo en particular y luego usar los llamados puntajes de factores como variables dependientes. Cuanto más relacionadas estén las variables unas con otras, menos factores se necesitarán para representar toda la *matriz* de variables.

Por ejemplo, digamos que usted está estudiando los efectos del conocimiento que los futuros padres tienen del sexo del bebé que está por nacer, sobre la forma como dichos padres perciben la personalidad del bebé. Un análisis factorial agrupa las variables similares de modo que algunas variables puedan representar un constructo en particular. A continuación, los investigadores asignan nombres a dichos grupos, o *factores*. La utilidad del análisis factorial radica en que permite a los investigadores examinar conjuntos de variables y ver qué tan íntimamente están relacionados entre ellos. Por ejemplo, en vez de manejar las variables contacto ocular, tacto y verbalización, podría manejarse un solo constructo llamado apego.

Significancia versus significado

¡Qué situación tan interesante para el investigador cuando descubre que los resultados de un experimento en realidad son estadísticamente significativos! Aunque usted esté apenas en los albores de su carrera, es probable que haya oído comentarios en su departamento y de otros estudiantes respecto a que el resultado absolutamente más deseable de cualquier investigación es que "los resultados sean significativos".

Lo que sus colegas quieren decir con esto, y lo que en realidad implica la significancia estadística, podrían ser dos cosas distintas. Lo que ellos quieren decir es que la investigación fue un *éxito técnico*, ya que la hipótesis nula no es una explicación razonable de lo que se observó. Ahora bien, si el diseño experimental y demás consideraciones se cuidaron debidamente, los resultados estadísticamente significativos son indudablemente el primer paso hacia hacer una contribución importante a la bibliografía en su campo. No obstante, hay que mantener en la debida perspectiva la presencia y la importancia de la significancia estadística.

Por ejemplo, tomemos el caso en que una muestra muy grande de adultos analfabetos (digamos, 10000) se divide en dos grupos. Un grupo recibe capacitación intensiva para leer utilizando computadoras, y el otro recibe capacitación intensiva para leer empleando enseñanza en un aula. El puntaje promedio del grupo 1 (que aprendió en el aula) es de 75.6 en una prueba de lectura, y es la variable dependiente. El puntaje promedio en la prueba de lectura para el grupo2 (que aprendió usando la computadora) es de 75.7. La varianza en ambos grupos es aproximadamente igual. Como puede verse, la diferencia en el puntaje es de sólo *un décimo* de punto (75.6 versus 75.7); sin embargo, cuando se aplica una prueba t para la significancia de la diferencia entre medias independientes, los resultados son significativos en el nivel de .01, lo que indica que las computadoras flincionan mejor que la enseñanza en el aula. En otras palabras, se minimiza el papel del azar. La diferencia de .01 en efecto es estadísticamente significativa, pero ¿significa algo? La mejoría en los puntajes de la prueba (por un margen tan pequeño) ¿es suficiente justificación para los \$300000 dólares que costaría equipar el programa con computadoras? ¿O dicha diferencia es tan insignificante que puede hacerse caso omiso de ella, aunque sea estadística-mente significativa?

He aquí otro ejemplo. Puesto que cuanto mayor es una muestra más próxima a las características de la población, a menudo basta una correlación muy pequeña entre dos variables para lograr la significancia estadística cuando el tamaño de la muestra es sustancial. Para 100 pares de puntajes, una correlación entre X y Y de .20 es significativa en el nivel de .05. El cuadrado de este coeficiente de correlación (o

coeficiente de determinación, que es un indicador de qué tan potente es la correlación) ¡sólo explica 4% (.2 al cuadrado) de la varianza! Esto implica que 96% de la varianza se queda sin explicación. Dada una relación estadísticamente significativa que no está ocurriendo exclusivamente por el azar, ¡eso es mucho que explicar! De hecho, si las muestras son suficientemente grandes, cualquier diferencia entre ellas será significativa.

A partir de estos dos ejemplos, podemos sacar algunas conclusiones acerca de la importancia de la significancia estadística.

Primera, la significancia estadística por sí sola no significa mucho a menos que el estudio realizado tenga una base conceptual sólida que confiera algún significado a la significancia del resultado.

Segunda, la significancia estadística no puede interpretarse independientemente del contexto en el que ocurre. Si usted es el superintendente de un sistema escolar, ¿está usted dispuesto a retener a los niños en el primer año escolar si el programa de retención aumenta significativamente sus puntajes en una prueba estandarizada medio punto (no la mitad del puntaje estándar)?

Por último, si bien la significancia estadística es importante como concepto, no es el objetivo final y ciertamente no debe ser la única meta de la investigación científica. Esa es la razón por la que nuestro propósito es *probar* hipótesis, no *demostrarlas*. Si nuestro estudio se diseñó correctamente, incluso los resultados negativos nos dicen algo muy importante. Si un tratamiento dado no funciona, esto es información importante que otros necesitan conocer. Si su estudio está bien diseñado, usted deberá saber por qué el tratamiento no funciona, y la siguiente persona que venga podrá diseñar su estudio teniendo en cuenta la valiosa información que usted proporcionó.

Metanálisis

Ya ha oído esto antes. Una de las características más importantes de la buena ciencia es que los resultados puedan repetirse. Por ejemplo, si usted utilizó con éxito cierta técnica para enseñar a leer a adultos analfabetos, sería bueno saber que la misma técnica puede usarse en circunstancias similares con una población similar; y los resultados serían los mismos.

Pero, ¿qué hay con el caso en que existen 10 o 500 incluso 100 estudios sobre el mismo fenómeno en los que se han usado diferentes números de sujetos, en diferentes circunstancias y hasta diferentes tratamientos o programas? Lo único que tales estudios tienen en común es el uso de la misma variable dependiente o de resultado, sea lectura, capacidad cognoscitiva, edad de aparición de la demencia senil, o cualquier otra de miles de variables dependientes posibles. ¿Cómo sacamos algo en claro de esta colección de hallazgos? ¿Es posible combinarlos, aunque las investigaciones que los produjeron difirieron unas de otras en muchos factores importantes como el tamaño y la selección de las muestras, las varia-bies de tratamiento, etcétera?

La respuesta es un sí condicionado. Gracias al **metanálisis**, una técnica relativamente nueva, es posible comparar los hallazgos de diversos estudios que tienen la misma variable dependiente. Antes de ver un ejemplo del funcionamiento del metanálisis, asegúrese de entender que "la misma variable dependiente" no implica necesariamente que se haya utilizado un instrumento idéntico en todos los estudios. Más bien, se midió la misma variable *conceptual* (como inteligencia o agresión o aprovechamiento). Si nos interesa estudiar un componente específico de la personalidad, podrían usarse diversos instrumentos, como el Cuestionario de 16 Factores de la Personalidad, o el Inventario Multifásico de la Personalidad de Minnesota, y los resultados de tales estudios se "combinarían" en un metanálisis.

Gene Glass acuñó el término metanálisis en 1976 para describir una estrategia de sin-tesis de los resultados de experimentos individuales. Se trata de un intento por integrar un cuerpo amplio y diverso de información acerca de un fenómeno en particular. Tenga presente que los datos para un estudio meta-analítico provienen de experimentos que ya se llevaron a cabo, no datos nuevos que todavía hay que recopilar y luego analizar. Efectivamente, ya se ha efectuado una buena parte del trabajo.

Cómo se realizan los estudios de metanálisis

He aquí un ejemplo de metanálisis realizado sobre la eficacia de los programas de intervención temprana (Castro y Mastropieri, 1988). Un metanálisis convencional tiene cuatro pasos, con un sinnúmero de variaciones en la forma en que se llevan a cabo.

Primero, se reúne el mayor número posible de estudios o un grupo lo más representativo posible de estudios sobre un fenómeno específico. Castro y Mastropieri utilizaron muchas de las técnicas y filentes que describimos en el capítulo 3 para averiguar qué estudios se habían realizado, incluidos *Dissertation Abstracts*, *ERIC* y *Psychological Abstracts*. Ellos también enviaron cartas a todos los investigadores que habían publicado en esta área o que habían participado en algún tipo de programa de intervención temprana. Castro y Matropieri se decidieron por 74 estudios, todos los cuales investigaron la efectividad de los programas de intervención temprana sobre niños de edad preescolar (con edades entre recién nacidos y 5 años) que tenían una discapacidad.

Segundo, es preciso convertir los resultados de los estudios a alguna métrica común para poder compararlos entre ellos. Esto es lógico, ya que de otro modo sería una pérdida de tiempo tratar de comparar cosas disímiles. La métrica empleada en muchos estudios meta-analíticos se denomina *tamaño del efecto*. Este valor se obtiene por medio de una comparación de las diferencias observadas entre los resultados para el grupo experimental (o el que recibió la intervención) y para el grupo de control (el que no recibió la intervención), medidas en alguna unidad estándar. Cuanto mayor es el tamaño del efecto, más grande es la diferencia entre los dos grupos. El empleo de la unidad estándar permite hacer comparaciones entre diferentes grupos y resultados, lo cual es el corazón de la técnica meta-analítica.

En un estudio meta-analítico, el tamaño del efecto se convierte en la variable dependiente. La variable independiente es el factor que se manipuló, sea el tipo de intervención, la edad de los niños, etc. En el estudio de Castro y Mastropieri, hubo un total de 215 comparaciones de grupo experimental-grupo de control, y un total de 215 tamaños de efecto de los 74 estudios que se reseñaron.

Tercero, los investigadores idean un sistema para codificar las diversas dimensiones del estudio, incluida una descripción de los sujetos, el tipo de intervención empleada, el diseño de investigación escogido, el tipo de resultado medido y las conclusiones a que llegaron los autores del estudio. A continuación, estos factores se utilizan en un examen de los tamaños de efecto calculados en el paso 2.

Por último, se aplican diversas técnicas descriptivas y correlacionales para examinar los resultados de los estudios como un todo. El investigador busca una tendencia o algo sustancial que tengan en común las direcciones de los resultados y que abarque todos los factores que se identificaron y codificaron tal como se describió en los dos pasos anteriores. En el estudio de Castro y Mastropieri, los investigadores concluyeron que los programas de intervención temprana sí producen "beneficios inmediatos moderadamente grandes en poblaciones con discapacidad". Estos beneficios al parecer aplican a resultados tales como puntajes de IQ, y aprovechamiento motor, del lenguaje y académico. No se observó eficacia del tratamiento para otros factores, como competencia social, concepto de uno mismo y relaciones familiares.

He aquí otro ejemplo para que usted vea el alcance de este tipo de estudios. Un estudio clásico examinó una pregunta clásica: ¿Funciona la psicoterapia? Smith y Glass (1977) realizaron un metanálisis de más de 375 estudios, que produjeron un total de 833 efectos. Estos 833 efectos representaron más de 25000 casos de sujetos experimentales y de control (los que recibieron y no recibieron psicoterapia). Un examen de los tamaños del efecto produjo diferencias marcadas y convincentes entre los grupos de participantes que recibieron terapia y los que no. Por otra parte, no hubo diferencias entre los tipos de terapia (como conductista o psicoanalítica).

¿Qué es lo que hace realmente atractiva esta técnica meta-analítica? En primer lugar, los estudios meta-analíticos hacen lo que hace la buena ciencia. Organizan datos y nos ayudan a entender qué significan. Imagine una lista de 375 estudios en la que los resultados de cada estudio se listan en una columna adyacente. ¡Qué difícil sería llegar a cualquier conclusión generalizable y válida! Para hacer las cosas todavía más confusas, digamos que algunos de los estudios utilizaron niños muy pequeños, otros usaron bebés, algunos examinaron habilidades sociales, otros inteligencia, etc. Podríamos tener un revoltijo de resultados. El metanálisis convierte esa maraña en algo comprensible.

Este capítulo fue una breve introducción al mundo de la estadística inferencial y a cómo el concepto de inferencia puede ofrecer herramientas muy potentes para tomar decisiones. En los dos últimos capítulos usted ha aprendido mucho acerca de cómo recopilar datos y luego examinarlos para detectar patrones, diferencias y relaciones. Ya está usted listo para explorar el primero de varios modelos de diseño empleados en los métodos de investigación; los métodos de investigación no experimentales.

Ejercicios

1. ¿Por qué el azar es inicialmente la explicación más atractiva de las diferencias observadas entre dos grupos?
2. Un investigador analizó los resultados de un experimento y observó que el valor t obtenido (en una prueba de medias independientes) era 1.29 con un total de 25 niños en el grupo 1 y 30 niños en el grupo 2. Utilice una tabla de valores críticos y diga si la hipótesis nula puede rechazarse o no.
3. ¿Cómo podría ser que los resultados de un estudio fueran estadísticamente significativos pero sin significado?
4. ¿Cómo funciona el teorema del límite central y por qué es tan importante para la estadística inferencial?
5. A partir del conjunto de puntajes siguiente, escoja una muestra aleatoria de 10 puntajes. Ahora repita el muestreo cuatro veces más hasta tener un total de cinco muestras distintas de tamaño 10.

51	5	5	5	
15	1	1	1	
45	5	5	4	
24	2	4	4	
12	4	5	4	
2	2	3	2	3

 - a. Calcule la media de toda la población.
 - b. Calcule la media de las medias
 - c. ¿Cómo podemos usar el teorema del límite central para explicar por qué las respuestas a los apartados a y b se parecen tanto?
 - d. ¿Cómo ilustra este ejemplo la potencia del teorema del límite central?

6. ¿Qué significa el término "estadísticamente significativo"?

7. Explique por qué un científico investigador no busca demostrar una hipótesis.
8. Como investigador; a usted le interesa el efecto de las guarderías de bebés en la firmeza del apego que surge entre el bebé y quien lo cuida. Usted sospecha que los bebés en las guarderías tendrán un apego poco firme a los 11 meses después de asistir a la guardería durante 2 meses, en comparación con los bebés que no asisten a la guardería, sino que son cuidados en su hogar por su principal cuidador. ¿Qué pasos generales seguiría usted para probar su hipótesis?
9. ¿Qué diferencia hay entre un error tipo I y uno tipo II?
10. Como investigador a usted le interesa examinar la efectividad de un nuevo programa de lectura. Usted planea hacerlo vigilando el avance en cuanto a habilidades de lectura de un grupo de jóvenes de primer año de bachillerato que usan el programa una vez por trimestre. ¿Cuál prueba estadística sería la más apropiada, y por qué?
 - a. prueba t para medias independientes
 - b. prueba t para medias dependientes
 - c. análisis de varianza
11. ¿Qué es un metanálisis, y por qué es importante como herramienta de investigación?

CAPITULO NUEVE

Métodos de investigación no experimentales

En Neil J. Salkind “Métodos de Investigación”, Prentice Hall, México, 1998.

Lo que aprenderá en este capítulo

- Cómo se realizan las investigaciones históricas, y en qué difieren de otros métodos
- Qué son las fuentes primarias y secundarias de datos, y cómo se usan
- Qué son la autenticidad y la exactitud de un estudio histórico, y por qué son importantes
- Cómo se usa la crítica interna y la externa para evaluar las investigaciones históricas
- Qué es la investigación por encuesta, y qué ventajas y desventajas tiene
- Qué tipos de preguntas contesta la investigación del desarrollo
- Las ventajas y desventajas de los métodos longitudinal y de corte transversal
- La importancia y el uso de los estudios de seguimiento
- El propósito y el uso de la investigación correlacional
- Cómo calcular e interpretar un coeficiente de correlación

En cierto sentido, los primeros ocho capítulos de este libro han servido como preparación para los tres siguientes, que se ocupan de tipos de diseños de investigación o métodos de investigación específicos. En este capítulo estudiaremos los métodos de investigación no experimentales, que son formas de examinar preguntas de investigación sin manipular directamente alguna variable.

Por ejemplo, si quisiera entender cómo ha evolucionado el uso de fármacos para tratar enfermedades mentales durante los últimos 100 años, utilizaría fuentes históricas. De forma similar, si quisiera conocer la frecuencia con que se utilizan castigos en diferentes niveles de seguridad en las prisiones, recopilaría información pertinente para contestar esa pregunta. Usted no manipularía ninguna variable (como colocar diferentes programas en un grupo de castigo alto o bajo). En el caso de la pregunta referente a las prisiones, usted estaría describiendo un resultado, y estaría efectuando investigación descriptiva.

Este capítulo se enfoca en las preguntas de investigación histórica y descriptiva, cómo se plantean y cómo se contestan. También estudiaremos preguntas de naturaleza correlacional que examinan la asociación entre variables y la importante distinción entre asociación (dos cosas que están relacionadas porque tienen algo en común) y causalidad (una cosa causa la otra).

Investigación histórica

Basta leer el prefacio del libro *Niñez victoriana* de Thomas Jordan (1987) para tener al menos una idea de lo diferente que es la investigación histórica de los tipos de investigaciones experimentales que por lo general vemos en las revistas científicas en el área de las ciencias sociales y del comportamiento. Los agradecimientos que Jordan expresa a las bibliotecas de la University of Missouri, Church of Jesus Christ of Latter-day Saints, Washington University; la Library of Congress, la British Library; la Royal Society of Health, el Reform Club y la Royal Statistical Society indican su contribución con datos de una forma u otra a su libro, que se enfoca hacia los niños de la era victoriana. Dichos datos, sean registros con 150 años de antigüedad de las estaturas de los niños o el porcentaje de niños menores de 15 años de edad que trabajaban en las plantas textiles (cerca de 14%), fueron sus "sujetos", y lo que hizo con ellos ejemplifica el meollo de nuestra explicación de la investigación histórica.

Niñez victoriana está organizado en nueve capítulos, cada uno de los cuales se concentra en un tema aparte, como ciudades, trabajo, vida

y muerte, aprendizaje y abogacía, y reforma. Jordan consultó fuentes de datos especialmente interesantes para apoyar conclusiones acerca de la forma en que se criaban y trataban los niños durante este periodo en Inglaterra. Los datos no son sólo tal o cual artículo de una revista escrito por otro erudito. Jordan a menudo recurrió a fuentes primarias (tenga paciencia, en breve explicaremos qué son) que tal vez no se nos hubiera ocurrido que existieran, mucho menos saber que están accesibles.

Nada más eche un vistazo a algunos de los materiales que usó:

- Datos de los registros de barcos que con regularidad transportaban niños "delincuentes" de Inglaterra a Australia.
- Poemas que reflejaban actitudes hacia los niños y el papel de los padres, como el siguiente:

El bebé

Si el bebé se toma las manos,
Y pide con sonidos y señas
Aquello que estás comiendo,
Aunque el corazón de madre se incline
Por darle lo que pide,
Recuerda, no puede masticar;
Y el alimento sólido le hace daño,
Aunque para ti sea muy bueno.

- Anuncios clasificados en los diarios, como el del número del 14 de septiembre de 1817 del *Morning Chronicle* que anuncia cuidados y educación para "JÓVENES DAMAS" quienes serán: "tratadas con la más tierna atención, estarán constantemente bajo su inspección inmediata y se formarán en todos los aspectos..." recibiendo instrucción sobre "Historia, geografía, escritura, aritmética y bordado...", todo por 30 guineas al año.

- El número de escuelas dominicales que abrieron sus puertas entre 1801 y 1851, por denominación (hubo al menos 11).

Es evidente que Jordan fue minucioso. Buscó aquí y allá y por todos lados hasta encontrar lo que necesitaba para presentar una imagen lo más completa posible de qué significaba ser un niño durante ese periodo. Al igual que todo buen científico, él recopiló datos (muy diversos y de una amplia variedad de fuentes) y organizó su información de manera tal que permitiera al lector llegar a ciertas conclusiones de las que no se hubiera percatado sin las labores de Jordan.

Realización de investigaciones históricas

La investigación histórica (**o historiografía**) en las ciencias sociales y de la conducta es considerada (injustamente) por algunos como investigación de segunda clase. La gente a menudo no puede decidir si tales investigaciones deben pertenecer a las ciencias sociales o a las humanidades, y con frecuencia termina siendo parte de algún dominio (historia de la educación, historia de la física, etc.) sin tener un hogar propio. No hay duda de que es una ciencia social, porque los historiadores reúnen y analizan datos lo mismo que cualquier científico social. Por otra parte, es una disciplina humanística también, porque los historiadores (o cualquier persona que realiza investigación histórica) examinan el papel que los individuos desempeñan en instituciones sociales como la escuela y la familia. Además, dado que a pocos científicos sociales y del comportamiento se les enseña acerca de la investigación histórica y su metodología correspondiente, pocos hacen realmente investigaciones en esa área o siquiera están familiarizados con la metodología apropiada. En general, son las personas llamadas "historiadores" las que se interesan en temas tales como la historia de las guarderías o de la reforma educativa o de los orígenes del psicoanálisis o en cientos de otros temas interesantes, y que finalmente son quienes efectúan las contribuciones importantes.

Entender la naturaleza histórica de un fenómeno es a menudo tan importante como entender el fenómeno mismo. ¿Por qué? Por la sencilla razón de que no podemos evaluar ni apreciar plenamente los avances logrados en la ciencia (sea psicología del desarrollo o física de partículas) sin entender un poco el *contexto* dentro del cual tuvieron lugar tales avances.

Por ejemplo, el desplazamiento de la población estadounidense hacia una mayor proporción de personas maduras y ancianas que ha ocurrido en los últimos 50 años es un suceso histórico que ha despertado un creciente interés en la gerontología como campo de estudio. Asimismo, entender las costumbres y condiciones de la era victoriana y la Viena de fines del siglo XIX (cuando Sigmund Freud inició el desarrollo de su teoría del psicoanálisis) proporciona ideas que nos ayudan a entender y apreciar mejor las teorías de Freud.

No es una afirmación ociosa aquella de "quienes no conozcan su pasado están condenados a repetirlo". Es verdad y es otra razón por la que debemos conocer el método histórico como parte de nuestro arsenal de habilidades en investigación.

Los pasos de la investigación histórica

Aunque es posible que a usted nunca le haya parecido así, realizar investigaciones históricas es en muchos sentidos muy similar a llevar a cabo cualquiera de los otros tipos de investigaciones que hemos mencionado en este libro.

Si bien es posible que los datos o la información básica difieran marcadamente, el investigador histórico sigue muchos de los mismos pasos que uno que utiliza cualquier otro método. Examinemos primero cada uno de esos seis pasos.

Primero, *los investigadores históricos definen un tema o problema* que desean investigar. La investigación histórica tiene un alcance ilimitado porque es un intercambio constante entre los sucesos de la actualidad y los del pasado. todo el pasado es la base de datos de los historiadores, una vasta colección de documentos e ideas, muchos de ellos difíciles de encontrar y más difíciles aún de verificar. Como una suerte de detectives, los investigadores históricos examinan de todo, desde las bitácoras de los barcos hasta los registros de nacimientos en las iglesias, para averiguar quién tiene parentesco con quién y qué papel ésta otra persona podría haber desempeñado en la comunidad. Es a partir de una inspección (que podría ser una simple lectura o una discusión con un colega) de este legado de información que a menudo surgen ideas para exploraciones subsecuentes.

Este paso es muy parecido a la labor mental de cualquier otro investigador, que por lo regular surge de un interés personal en un área dada. Por ejemplo, podría interesarnos la historia de la reforma educativa y específicamente la noción del origen de las leyes que obligan a los niños a asistir a la escuela.

Segundo, hasta donde es posible, *el investigador formula una hipótesis*, que por lo regular se expresa como pregunta. Por ejemplo, la pregunta general podría ser "¿Cuándo, como y por qué la escuela se volvió obligatoria para los niños de menos de 16 años? Aunque plantear hipótesis en una forma no declarativa es algo que no suele hacerse en los estudios científicos, la investigación histórica exige adoptar un conjunto distinto de reglas. Algunos de los criterios para caracterizar como buena una hipótesis, que tratamos en el capítulo 2, se pueden aplicar a la investigación histórica (como el que una hipótesis es una conjetura educada), pero otros no (como el de buscar relaciones estadísticas entre variables).

Tercero, al igual que en cualquier otra empresa de investigación, es preciso *utilizar diversas fuentes para reunir datos*. Como veremos en breve, estas fuentes difieren marcadamente de aquellas a las que usted probablemente está acostumbrado. Si bien las entrevistas pueden ser una fuente de datos en casi cualquier tipo de investigación, el análisis de documentos escritos y la extracción de datos de expedientes y demás suele estar dentro del ámbito del investigador histórico.

Cuarto, es preciso evaluar los indicios tanto en cuanto a su *autenticidad* como a su *exactitud*. Hablaremos más de esto posteriormente.

Quinto, *es necesario sintetizar o integrar los datos* para obtener un cuerpo de información coherente. Esto es similar a los pasos que usted tal vez siguió cuando revisó la bibliografía durante la preparación de una propuesta, pero en este caso lo que se integran son resultados, y lo que se busca son tendencias o patrones que podrían sugerir otras preguntas que valga la pena hacerse.

Por último, al igual que en cualquier otro proyecto de investigación, hay que *interpretar los resultados* a la luz del argumento que se presentó originalmente para justificar la investigación del tema, y a la luz de la pregunta que se hizo cuando se inició la investigación. Su habilidad como intérprete tendrá mucho que ver con lo bien preparado que esté para entender los resultados de su recopilación de datos. Por ejemplo, cuanto más sepa usted acerca del clima económico, político y social de fines del siglo XIX y principios del siglo XX, mejor podrá entender cómo, por qué y cuándo la asistencia obligatoria a la escuela se convirtió en la regla más que la excepción.

Fuentes de datos históricos

Los historiadores por lo regular se apoyan en dos tipos de datos: fuentes primarias y fuentes secundarias. Cada uno desempeña un papel específico en la realización de las investigaciones históricas, y son igualmente valiosos.

Fuentes primarias de datos históricos

Las **fuentes primarias** de datos históricos son objetos, documentos, entrevistas y registros de testigos oculares, historias orales, diarios y expedientes escolares *originales*. Por ejemplo, si usted quiere saber cómo las familias japonesas que vivían en Estados Unidos durante la Segunda Guerra Mundial se ajustaron a la política de internarlas en campos especiales, la entrevista que usted haga a un hijo de una de esas familias sería una fuente primaria. Lo mismo puede decirse de un diario llevado por un adulto en el que asentó sus experiencias.

Las fuentes primarias son los resultados directos de un suceso o una experiencia y que se registran sin que el historiador necesariamente tenga la intención de utilizar posteriormente la referencia. Tales fuentes podrían ser un noticiario que se exhibió en un cine hace 50 años o un registro del número de personas que recibieron psicoterapia en 1952 o las minutas de una reunión de una junta escolar. Si usted filera un historiador lo único que le impediría formarse una imagen muy exacta de qué se sentía estar presente en esa reunión de la junta escolar es el hecho de que está percibiendo el punto de vista de otra persona a través de las minutas. No obstante, usted estaría tan cerca de estar ahí como es posible.

En la tabla 9.1 se resumen algunas fuentes de datos primarias.

Fuentes secundarias de datos históricos

En tanto las fuentes primarias son relatos de primera mano de los sucesos, las fuentes secundarias de datos históricos son de segunda mano o al menos están alejadas un paso del suceso original, como un resumen de estadísticas importantes, una lista de fuentes primarias

importantes y una columna periodística basada en un relato de un testigo ocular (el relato mismo sería una fuente primaria). Estas fuentes relatan lo que otros presenciaron, digamos un transeúnte, pero que la fuente no presenció directamente. Y tal como sucede en el juego del "teléfono descompuesto", a menudo algo se pierde en la traducción.

La consideración más importante al usar fuentes secundarias es qué tanto puede confiarse en la fuente original de los datos. Por ejemplo, un nuevo análisis de los datos que Sir Cyril Burt obtuvo sobre gemelos hace 100 años llevó a varios científicos a concluir que los datos habían sido falsificados. Una buena parte de lo que se conoce (y que se creía verdad) acerca de la naturaleza de la inteligencia, por ejemplo, se basó en el análisis inicial.

Fuentes primarias o secundarias: ¿A cuáles dirigirse?

Sería un mundo ideal para el historiador si las fuentes primarias siempre estuvieran disponibles, pero es común que no sea así. Al igual que en tantas otras situaciones del mundo de la investigación, el ideal (como la muestra perfecta) simplemente no es asequible. En vez de ello, hay que conformarse con lo mejor que se encuentre, que podría ser una fuente secundaria.

(1) Categoría	Ejemplo
Documentos	minutas de reuniones contratos escrituras testamentos permisos fotografías listas cuentas películas catálogos mapas relatos periodísticos que están narrados en primera persona diarios registros de graduación
Historias orales	relatos de sucesos, hablados o asentados transcripciones de tribunales
Restos, residuos y reliquias	herramientas alimentos objetos religiosos ropa construcciones equipo libros pergaminos

Tabla 9.1 Lista de algunas fuentes primarias de datos históricos. ¿Puede usted imaginar el trabajo que se requiere para localizar algunas de estas fuentes?

Dado que ambos tipos de fuentes pueden ser igualmente útiles (y confiables), los investigadores no deben preferir implícitamente un tipo al otro, ya que los dos proporcionan información valiosa. Por ejemplo, sería un tanto difícil entrevistar a los maestros que enseñaban en la Inglaterra victoriana que Jordan describió, pero podríamos darnos una buena idea de qué sucedía durante un día de escuela leyendo una carta de un padre al director de la escuela. Los buenos investigadores no lamentan la falta de fuentes primarias o si falta de una u otra carta potencialmente importante. En vez de ello, sacan el mayor provecho posible a lo que tienen.

He aquí otro ejemplo. Para quienes están interesados en el desarrollo infantil, existe un increíble depósito de manuscritos y materiales visuales en Antioch College en Yellow Springs, Ohio, donde es posible encontrar ambos tipos de fuentes. Ahí, la *Society for Research in Child Development* ha almacenado (y sigue solicitando) miles de fuentes primarias y secundarias relacionadas con los niños y sus familias, en muchos casos contribuidas por los científicos que realizaron originalmente los trabajos. Entre los materiales que se pueden encontrar ahí están:

- correspondencia entre investigadores acerca de un tema específico,
- cartas personales que incluyen información acerca de ideas y avances hacia una meta específica.

- borradores de lo que después serían artículos de investigación importantes,
- datos originales que otras personas interesadas en la misma área pueden utilizar y analizar con nuevas técnicas,
- películas de estudios de investigación, como las que detallan el crecimiento y desarrollo de niños pequeños compiladas por "edades y etapas" por el doctor Arnold Gesell, y
- programas y órdenes del día de cientos de reuniones de sociedades profesionales enfocadas en los niños.

Autenticidad y exactitud

Así como los investigadores que utilizan pruebas de aprovechamiento como fuente de datos tienen que asegurarse de que la prueba sea confiable y válida, los historiadores necesitan establecer el valor de los datos de fuentes primarias y secundarias en los que se fundamentan sus argumentos. Al igual que otros, los historiadores necesitan adoptar una actitud crítica y evaluadora hacia la información que recaban. De lo contrario, el documento primario inexacto de hoy (tal vez una falsificación) se convierte mañana en la fuente de información de otro historiador. El ciclo se repite, y la fuente primaria de uno se convierte en la fuente secundaria de otro, y toda la base de datos se contamina cada vez más con información no auténtica.

La evaluación de los datos primarios y secundarios se efectúa aplicando dos criterios distintos: **autenticidad** (también conocida como crítica externa) y **exactitud** (también llamada crítica interna).

La crítica externa como criterio

La **crítica externa** aplicada a datos históricos se ocupa de la *autenticidad* de los datos. Básicamente, este criterio pregunta si los datos son genuinos y confiables o si son falsos. ¿Se escribieron cuando se dice que se escribieron? ¿Los escribió la persona que los firma? ¿Se hallaron donde se dejaron? Éstas son sólo algunas de las preguntas que debemos hacer antes de poder confiar en los datos.

La autenticidad de un documento u otra fuente primaria a veces es fácil de establecer; pero otras veces es poco menos que imposible. Se puede examinar; por ejemplo, la antigüedad y la calidad de tintas específicas para fechar un documento. Los tipos de estilos de escritura, las técnicas de imprenta, las composiciones químicas del papel, el uso del lenguaje y los conocimientos generales son indicadores de cuándo (e incluso cómo) se preparó un documento. Lo principal que el historiógrafo necesita buscar es la consistencia. ¿Embonan todas las piezas como en un rompecabezas o hay datos extremos importantes que simplemente no ajustan y hacen surgir dudas? ¿Y qué valor puede tener cualquier trabajo que se realice si los datos en los que se basa son dudosos?

Por ejemplo, la presencia de monedas antiguas en los mismos recipientes que contenían los famosos pergaminos del Mar Muerto aportó pruebas adicionales de que los pergaminos eran tan antiguos como se sospechaba. Las monedas y algunas pruebas forenses muy avanzadas, como el fechado con carbono, llevaron a la conclusión de que los pergaminos tenían unos 2,500 años de antigüedad (al momento de escribirse este libro).

El historiador principiante no cuenta con la capacitación ni las técnicas necesarias para realizar análisis tan sofisticados, por lo que se ve obligado en mayor o menor grado a basar sus decisiones respecto a la autenticidad en diversos indicios y emitir un juicio acerca de la utilidad de los datos. Incluso si usted no posee las herramientas, debe cerciorarse de haber agotado todas las posibilidades de establecer la autenticidad de sus datos. De lo contrario, sus labores de investigación podrían ser en vano.

La crítica interna como criterio

Un segundo criterio de evaluación es la **crítica interna**, que se ocupa de la *exactitud*, o de qué tan confiable es la fuente como un reflejo verdadero de lo que sucedió. ¿Son creíbles las cifras de la encuesta realizada en 1890 para determinar cuántos niños asistían a la escuela? ¿Los informes hechos por los padres acerca de los cambios repentinos de ánimo de los adolescentes en la década de 1950 son un reflejo exacto del comportamiento real de los jóvenes?

Una forma de determinar el nivel de exactitud es pedir a un experto que examine documentos o reliquias y emita su opinión en cuanto a si es o no un reflejo exacto de los sucesos ocurridos durante el periodo que se investiga.

Las limitaciones de la investigación histórica

No hay duda de que la investigación histórica tiene algunas deficiencias importantes en comparación con otros métodos de investigación en las ciencias sociales y de la conducta.

En primer lugar, puesto que la disponibilidad de los datos siempre está limitada por factores que no están bajo el control del investigador, *es probable que la generalizabilidad de los resultados también sea limitada*. Si todo lo que tenemos para trabajar son cartas, sin algo que verifique si los sucesos realmente ocurrieron, no podemos tomar mucho de los hallazgos obtenidos y aplicarlo a otra época o situación. De hecho, los historiadores a menudo tienen que conformarse con lo que pueden obtener para estudiar un tema dado, en vez de lo ideal.

Segundo, *los datos de las investigaciones históricas a menudo se ponen en duda porque se derivan primordialmente de las observaciones de otros*, sean cartas, libros u obras de arte. Quienes piensan que la observación de primera mano (pruebas, tareas, etc.) produce

información con el mayor potencial para entender el comportamiento pueden tener razón en parte, pero eso no justifica hacer caso omiso de los otros tipos de datos que la historia ofrece.

Tercero, las investigaciones históricas a menudo son *una tarea larga y ardua que puede incluir cientos o miles de horas* de examinar documentos (¡Si es que puede encontrarlos!) en busca de pistas e indicios que puedan apoyar la hipótesis.

Cuarto, dado que los criterios que normalmente se aplicarían a las investigaciones empíricas incluyen factores como la confiabilidad y validez de los instrumentos empleados, en la investigación histórica *se usan otros criterios menos rigurosos (pero mas exhaustivos) para evaluar las herramientas de medición.*

Investigación descriptiva

Aunque son varios los factores que distinguen los diferentes tipos de investigación, es probable que el factor más importante sea el tipo de pregunta que se desea contestar. Si estamos realizando investigación histórica, estamos tratando de entender sucesos que ocurrieron en el pasado y la relación que podrían tener con los sucesos actuales. Generamos preguntas o hipótesis, recabamos datos y continuamos como si estuviéramos realizando cualquier tipo de investigación.

La **investigación descriptiva** es diferente. El propósito de la investigación descriptiva es *describir la situación prevaleciente en el momento de realizarse el estudio.* Por ejemplo, si queremos saber cuántos maestros utilizan un método de enseñanza dado, podríamos pedir a un grupo de estudiantes que llenen un cuestionario, con lo cual mediríamos el resultado en el momento de ocurrir. Si quisiéramos saber si hay diferencias en la frecuencia de uso de ciertos tipos de palabras entre niños de 3, 5 y 7 años, *describimos* dichas diferencias dentro de un marco descriptivo o de desarrollo.

Al igual que con la investigación histórica, la diferencia más significativa entre la investigación descriptiva y la investigación causal comparativa o experimental (que mencionamos en el capítulo 1 y trataremos con detalle en el capítulo siguiente) es que la investigación descriptiva no incluye un grupo de tratamiento ni uno de control. No estamos tratando de mostrar la influencia de una variable sobre otra. En otras palabras, lo único que estamos haciendo por el lector del informe de investigación es pintar una imagen. Cuando alguien lee un informe que incluye uno de los varios métodos descriptivos que veremos, deberá poder ver una imagen mayor de lo que sucedió. Si bien podría haber cabida para la pregunta *¿por qué?*, ésta generalmente se deja para un enfoque más experimental.

Aunque hay muchos tipos de investigaciones descriptivas, nuestra explicación se concentrará en

- estudios de casos o estudios a fondo de personas u organizaciones individuales
- estudios de desarrollo en los que se examinan cambios o diferencias en el desarrollo, y
- estudios correlacionales, en los que se describen relaciones entre variables.

Estudios de casos

Hubo una vez una niña llamada Genie que estuvo aislada de otros seres humanos durante la mayor parte de sus primeros años (Curtiss, 1977). Cuando por fin fue descubierta y puesta en libertad a los 14 años de edad, proporcionó a los psicólogos copiosa información acerca de los efectos del retardo en hablar sobre el desarrollo del lenguaje.

Psicólogos y lingüistas estudiaron su desarrollo en lo tocante al lenguaje utilizando un **estudio de caso**. Un estudio de caso es un método empleado para estudiar un individuo o una institución en un entorno o situación único y de una forma lo más intensa y detallada posible. La palabra único es crítica aquí porque el investigador está tan interesado en las condiciones existentes que rodean a la persona como en la persona misma. Es la calidad de unicidad lo que separa a esa persona (y al caso) de otras.

Es posible que usted haya oído ya el término *estudio de caso*. La idea del estudio de caso representa una parte importante de la metodología que los médicos usan para recabar y diseminar información. Las revistas médicas publican con regularidad estudios de casos de individuos cuyas condiciones son tan inusitadas que sus síntomas y tratamiento exigen atención especial, y es necesario diseminar información acerca de sus casos.

Un pionero en el uso de los estudios de casos fue el médico convertido en psicólogo Sigmund Freud, quien basó en ellos su teoría psicoanalítica del desarrollo de la personalidad. Su famosa paciente, Anna O., y sus observaciones detalladas acerca de su condición condujeron al uso de la asociación libre como método para el tratamiento de la histeria y otras condiciones. También tenemos el trabajo de Jean Marc Itard, uno de los primeros "educadores especiales" y su descripción como estudio de caso del niño salvaje de Avyron, en quien se basó la popular película *El niño salvaje*.

Los estudios de casos no están limitados a las personas. La Harvard Business School tiene como práctica regular incluir estudios de casos de negocios que fracasan o tienen éxito como parte fundamental del programa de estudios de sus estudiantes de posgrado. La investigación de un caso, bajo el microscopio por así decirlo, permite a los estudiantes repasar los pasos seguidos y entender mejor la mecánica de cómo un negocio podría resultar afectado por diversos factores. También las familias, escuelas, pandillas y organizaciones sociales son grupos legítimos para el enfoque de estudio de caso.

Por ejemplo, la conocida descripción de una escuela experimental, *Summerhill* (Neil 1960), es un estudio de caso complejo y detallado de una escuela inglesa única basada en la idea de una educación "abierta". Un trabajo similar, más reciente, es *Among School Children*, de Tracey Kidder, un estudio de caso narrado de una maestra de quinto año y sus actividades a lo largo de un año escolar. En parte gracias a la habilidad de estos escritores y en parte gracias a la estructura de estudio de caso de los libros, el lector obtiene una visión íntima de la vida en esos dos diferentes tipos de escuelas. Y no debemos olvidar al autor Jonathan Kozol, quien en sus libros *Rachael and Her Chudren*, *Savage Inequalities* y *Amazing Grace* informó a la comunidad social más amplia cómo las escuelas deficientes, la falta de hogar y la pobreza afectan a niños y familias individuales.

Algunas ventajas del método de estudio de caso

Los estudios de caso son una forma única de capturar información acerca de la conducta humana por varias razones.

Primera, los estudios de caso se enfocan hacia un solo individuo o cosa (sea, por ejemplo, una persona o un distrito escolar), lo que permite *un examen y escrutinio muy de cerca y la recopilación de una gran cantidad de datos detallados*. Es por estas razones que los estudios de caso siempre han sido populares como método para situaciones clínicas.

Segunda, los estudios de caso *fomentan el uso de varias técnicas distintas para obtener la información necesaria*, las cuales van desde las observaciones personales hasta las entrevistas de otras personas que podrían conocer el objetivo del estudio de caso hasta los expedientes de escuelas o doctores relacionados con la salud y otras cuestiones.

Tercera, *sencillamente no hay mejor manera de obtener una imagen más completa de lo que está ocurriendo* que a través de un estudio de caso. Esto es exactamente lo que Freud hizo en sus primeros trabajos. Ciertamente, él no podía haber usado un cuestionario para preguntar acerca de los sueños de sus pacientes, ni era concebible alcanzar su nivel de análisis utilizando otra cosa que no fuera un escrutinio intensivo de los detalles aparentemente más insignificantes de la forma como la mente funciona. Estos datos contribuyeron a su extraordinario conocimiento del funcionamiento de la mente humana y a la primera teoría aceptada de las etapas del desarrollo humano.

Cuarta, si bien los estudios de casos no prueban hipótesis, *sugieren direcciones para estudios subsecuentes*.

Algunas desventajas del método de estudio de caso

Si bien el estudio de caso ha proporcionado información muy importante (que probablemente no podría haberse revelado de alguna otra manera) sí tiene sus desventajas.

Primera, como sucede con tantas otras cosas, *lo que se ve no es siempre lo que se obtiene*. La realización del estudio de caso podría parecer fácil (sólo hay que encontrar un sujeto, una escuela, un aula, una oficina, una familia, etc.), pero de hecho es el método de investigación que consume más tiempo del que pudiera imaginarse. Es necesario recopilar datos de una amplia variedad de situaciones y fuentes, en muy variadas condiciones, y dichas situaciones y condiciones casi nunca pueden escogerse. Si el niño que usted está observando se queda en el salón en vez de salir al recreo, usted también tendrá que hacerlo.

Segunda, *las notas que usted asienta en su bitácora o diario podran reflejar con exactitud la realidad (o lo que usted observa), pero también podrían no hacerlo*. Todo mundo aborda cualquier situación con cierta predisposición, y los investigadores deben procurar que esa predisposición no interfiera el proceso de recopilación e interpretación de los datos. Un paso en la dirección correcta aquí es reconocer que usted está predispuesto (lo mismo que yo o que su mejor amigo), para poder asegurarse de que las conclusiones que saque no estén basadas en una realidad predispuesta.

Tercera, *lo que los estudios de caso proporcionan en profundidad, pierden en amplitud*. Aunque están extremadamente enfocados, no abarcan tanto como otros métodos de investigación. Por ello, los estudios de caso sólo constituyen el método recomendado si usted desea efectuar un estudio a fondo de un tipo de fenómeno.

Cuarta, *ni siquiera se le ocurra tratar de establecer relaciones de causa y efecto entre lo que ve y lo que usted piensa que podría estar causando los resultados*. Aunque se sienta tentado a especular, la estrategia de estudio de caso no permite en absoluto llegar a tales conclusiones. No sólo son insuficientes los datos (*n vale 1*) para concluir que existe una relación de causa y efecto, sino que, lo cual es más importante, estudiar relaciones causales no es el propósito del método. Si quiere estudiar relaciones causales, tendrá que usar herramientas cuya idoneidad para ese propósito sea generalmente aceptada.

Por último, por la naturaleza misma de los estudios de casos, *la generalizabilidad de los hallazgos es limitada*. Aunque podríamos poder aprender algo acerca de otro niño u otra institución similar a la que se estudió, no es prudente concluir que por ser similar el foco del estudio, los hallazgos también han de ser similares.

Algunos científicos piensan que los estudios de casos nunca tendrán como resultado investigación básica de avanzada (lo cual de todos modos no es su propósito). La utilización de los estudios de caso es revelar una diversidad y riqueza de conducta humana que sencillamente no está accesible por ningún otro método.

Investigación por encuesta

La mejor aplicación de la teoría y la práctica del muestreo probablemente pueda encontrarse en la investigación por encuesta. Los investigadores por encuesta intentan estudiar directamente las características de las poblaciones realizando encuestas. Tal vez el tipo de encuestas con el que usted está más familiarizado es el que suele llevarse a cabo cuando hay elecciones. En esas encuestas se hacen preguntas a muestras relativamente pequeñas de votantes en potencia (unos 1200) acerca de sus intenciones de voto en las elecciones. Para crédito de los diseñadores de las encuestas, los resultados a menudo se acercan mucho a los resultados reales después de la elección.

Las investigaciones por encuesta, también llamadas encuestas de muestreo, examinan la frecuencia y las relaciones entre variables psicológicas y sociológicas e indagan sobre constructos como actitudes, creencias, prejuicios, preferencias y opiniones. Por ejemplo, se podría usar una encuesta de muestreo para determinar

- las actitudes de los padres hacia la imposición de castigos en las escuelas,
- las actitudes de los residentes del vecindario hacia nuevas restricciones de estacionamiento,
- la percepción que tienen los adolescentes sobre el establecimiento de una hora límite para andar en la calle, y
- el uso de drogas en las escuelas de educación media.

La entrevista

La herramienta básica que se utiliza en la investigación por encuesta es la entrevista. Las entrevistas (o cuestionarios orales) pueden adoptar varias formas, desde una sesión de preguntas y respuestas totalmente informal realizada en la calle hasta una interacción altamente estructurada y detallada. De hecho, muchos de los puntos que se citaron al describir los cuestionarios aplican también a las entrevistas. Por ejemplo, aunque no es necesario preocuparse por el formato físico de las preguntas en una entrevista (ya que el entrevistado nunca las ve), sí hay que ocuparse de cuestiones como las transiciones entre secciones, ser sensibles al tipo de información que se está solicitando, y ser objetivos y directos.

La mayor parte de las entrevistas inicia con lo que se denomina **información de hoja de portada**. Se trata de la información neutral acerca del entrevistado, como su edad, dónde vive, cuántos hijos tiene, sus ingresos, género y nivel educativo. Esta información ayuda al entrevistador a lograr varias cosas.

Primero, ayuda a establecer una relación de comunicación entre el entrevistador y el entrevistado. Las preguntas como: "¿a qué universidad asistió?" o "¿Cuántos hijos tiene?" podrían no ser intimidantes.

Segundo, establece un conjunto de datos que caracteriza a la persona que se está entrevistado. Tales datos pueden ser valiosísimos en el análisis del foco principal de la entrevista que se efectúa más adelante en la encuesta.

Hay dos tipos generales de preguntas que se incluyen en las entrevistas. Las preguntas estructuradas o cerradas tienen un foco claro y evidente y una respuesta que es obviamente la que se pide. Tales preguntas son tan claras para el entrevistador como para el entrevistado. Preguntas como: "¿a qué edad comenzó usted a filmar?" y "¿Cuántas veces ha visitado usted esta tienda?" piden respuestas explícitas. Por otro lado, las preguntas no estructuradas o abiertas permiten al entrevistado proporcionar más o menos detalles en sus respuestas. Preguntas como "¿Por qué se opondría usted a la legalización de las drogas que ahora están prohibidas?" o "¿Cómo resolvería usted el problema de los embarazos en la adolescencia?" permiten una respuesta más amplia por parte del entrevistado. En ambos casos, el entrevistador puede hacer preguntas adicionales sobre el tema.

Las entrevistas pueden ser muy útiles si usted desea obtener información que de otra forma sería inaccesible, lo que incluye el conocimiento de primera mano de los sentimientos y percepciones de las personas. Por ejemplo, en un estudio realizado por Smith y Shepard (1988), se entrevistó a maestros y padres como parte de una estrategia multifacética para entender cuándo los niños están listos para asistir al jardín de niños y permanecer en él. En el estudio, las entrevistas se combinaron con otras técnicas como observaciones en clase y el análisis de documentos importantes. Los investigadores sacaron mucho provecho a los resultados de las entrevistas al examinar estos resultados a la luz del resto de la información que habían recopilado durante el estudio.

En el aspecto positivo, las entrevistas ofrecen gran flexibilidad porque las preguntas se pueden guiar en cualquier dirección (dentro del ámbito del proyecto). También se puede observar la conducta no verbal del entrevistado, la situación y otra información que podría ser valiosa. Otra ventaja de las entrevistas es que es posible establecer el tono general y la orientación de la entrevista según convenga al entrevistador (hasta cierto punto, desde luego).

Por otra parte, las entrevistas también tienen un lado negativo. Consumen mucho tiempo, el cual es costoso. Entrevistar 10 personas podría tomar 20030 horas, incluido el tiempo para desplazarse y cosas así. Además, al no disfrutar del anonimato total que ofrecen, por ejemplo, los cuestionarios, los entrevistados podrían estar reacios a responder de manera tan honesta como lo harían en el anonimato. Otras desventajas son la predisposición propia del entrevistador y la falta de un conjunto estandarizado de preguntas. Un buen entrevistador sondeará a fondo al entrevistado para obtener información adicional, tal vez de diferente índole que la que buscaría otro entrevistador que comenzara con las mismas preguntas. Aunque hacer preguntas de seguimiento es una práctica excelente, ¿qué se hace con la entrevista en la que el sondeo no condujo a la misma información y por tanto a los mismos resultados?

Como diseñar una entrevista

El diseño de una entrevista se inicia de manera parecida al diseño de cualquier propuesta para un proyecto de investigación. El primer paso es plantear el propósito de la entrevista teniendo en cuenta las metas del proyecto. Luego, como siempre, se revisa la bibliografía pertinente para averiguar qué se ha hecho en el pasado y si se han realizado o no otros estudios por entrevista. Usted podría incluso encontrar la entrevista que se usó y aprovechar partes de ella en su propia investigación. Ésta es una práctica muy común cuando los entrevistadores emplean la misma entrevista, digamos, 10 años después para determinar si ha habido cambios en las tendencias.

Segundo, seleccione una muestra que sea apropiada para su estudio tanto en sus características como en su tamaño. Si quiere conocer los sentimientos hacia los disturbios raciales en Estados Unidos, no se debe entrevistar únicamente a ciudadanos de raza blanca; hay que incluir a todas las minorías. Asimismo, a pesar de que las entrevistas requieran mucho tiempo y esfuerzo, no es posible escatimar el tamaño de la muestra pensando que lo que se pierde en tamaño se puede compensar en riqueza y detalle. Así no es como funciona.

El siguiente paso es *crear las preguntas*. Ya sabemos que las preguntas, sean estructuradas o no estructuradas, deben ser claras y directas, sin intenciones ocultas, dobles negativos, palabras rebuscadas difíciles de entender, etc. Una de las mejores formas de averiguar qué tan buena es una entrevista, es sometiéndola a una "prueba de campo". Utilícela con personas que tengan las mismas características que el público al que va dirigida. Solicite retroalimentación y haga los cambios que considere pertinentes.

Una vez que el formato de la entrevista está (más o menos) definido, llega el momento de *capacitar a los entrevistadores*. La mayor parte de los rasgos que se buscan en un entrevistador son obvios: amable, buena presentación, aspecto poco provocativo y lo bastante responsable como para llegar puntual al sitio de la entrevista. Sin embargo, estas cualidades a menudo no son suficientes. Los entrevistadores necesitan aprender a ir más allá de la pregunta si es necesario. Por ejemplo, si usted está haciendo preguntas acerca de disturbios raciales, el entrevistado podría decir

"Si. A veces siento que me discriminan.»

Si usted no pregunta "¿Por qué?" y da seguimiento a la respuesta del entrevistado, perderá información que podría ser valiosa e interesante. La mejor manera de capacitar entrevistadores es hacer que un entrevistador avezado observe mientras los candidatos entrevistan a un sujeto de práctica y haga comentarios.

Por último, llega el momento de realizar las entrevistas reales. *Destine a ello abundante tiempo* y ponga manos a la obra. No sea tímido, pero tampoco sea demasiado agresivo.

Los diez mandamientos de las entrevistas

Si se ha esforzado en prepararse para la entrevista, no deberá toparse con problemas importantes. No obstante, hay ciertas cosas que debe tener presente para llevar a cabo su entrevista con mayor facilidad y sacar más provecho de ella posteriormente, en el momento de examinar los resultados de su labor.

Con esto en mente, he aquí los 10 mandamientos de las entrevistas (fanfarrias, por favor). No olvide que muchos de éstos, si no todos, podrían también clasificarse como efectos del entrevistador, ya que el comportamiento del entrevistador puede afectar de manera significativa el resultado.

1. No inicie la entrevista en frío. Entre en calor con algo de conversación sobre cualquier tema, desde el estado del tiempo hasta la final del fútbol (sobre todo si el juego es esa noche y usted sabe que el entrevistado es aficionado a ese deporte). Utilice cualquier cosa que pueda ayudarle a romper el hielo y hacer más cálida la interacción. Si le ofrecen café, acéptelo (aunque no se lo beba todo si no quiere). Si no le gusta el café, rechácelo amablemente o pida un sustituto.
2. Recuerde que su misión es obtener información. Concéntrese en su tarea y utilice una serie de preguntas impresas como ayuda.
3. Sea directo. Conozca sus preguntas lo suficiente como para no tener que estar consultando continuamente su hoja, pero no dé la impresión de ser casual o carente de interés.
4. Vista de manera apropiada. Quitele cinco de sus seis arracadas, si piensa que llevar seis podría predisponer al entrevistado.
5. Encuentre un lugar tranquilo para realizar la entrevista, donde usted y su entrevistado no estén sujetos a distracciones. Cuando haga la cita para la entrevista, decida en qué lugar se hará. Si un sitio propuesto no es aceptable (como "en la cafetería"), sugiera otro (como el vestíbulo de la biblioteca). Llame la víspera de su entrevista para confirmar la visita. ¡Se sorprendería al saber cuántos entrevistados se olvidan de la cita!
6. Si su entrevistado no le proporciona una respuesta satisfactoria la primera vez que usted hace una pregunta, plantéela de otra manera. Siga replanteándola parcial o totalmente hasta acercarse lo más posible a lo que usted cree necesitar.
7. Tal vez usted quiera usar una grabadora para poder repasar las respuestas después. Si lo hace, debe tener en cuenta varias cosas. Primera, no olvide pedir permiso para grabar la entrevista antes de comenzar. Segunda, la grabadora no debe ser una "muleta". No deje que la cinta corra sin tomar notas ni obtener toda la información que pueda durante la entrevista misma.
8. Haga que el entrevistado se sienta como una parte importante de un proyecto importante y no sólo alguien que está presentando una prueba. A la mayoría de la gente le gusta hablar si se le da la oportunidad. Dígales que se da cuenta de lo valioso que es su tiempo y lo mucho que aprecia su participación. ¡No olvide prometer y entregar, en su momento, una copia de los resultados!
9. Para convertirse en un buen entrevistador se requiere lo mismo que para convertirse en un concertista: practicar, practicar y practicar. Su primera entrevista, como la de todo mundo, estará llena de aprensiones y dudas. A medida que haga más entrevistas, su

mayor confianza y dominio de las preguntas producirán una interacción menos tensa y usted obtendrá información más útil.

10. Dé las gracias al entrevistado, y pregúntele si tiene dudas.

Otros tipos de encuestas

¿Alguna vez ha estado usted en casa a la hora de la cena y el teléfono suena, y la persona que llama le pregunta con qué frecuencia usted toma el camión, recicla el periódico, utiliza una computadora o renta un automóvil?

Esas llamadas representan uno de varios tipos de investigación por encuesta, todas las cuales son de naturaleza descriptiva. Además de las entrevistas, que son el método de investigación por encuesta primario, y las encuestas telefónicas, están los paneles o grupos de enfoque (donde se entrevista y vuelve a entrevistar a un grupo pequeño de personas) y los cuestionarios por correo.

Cómo realizar la investigación por encuesta

La investigación por encuesta parte de un plan general, llamado plan de flujo, que indica cuáles actividades ocurrirán cuándo. El plan comienza con el objetivo del estudio, sigue con los diversos métodos que podrían utilizarse para recopilar los datos y termina con un informe final y un resumen de los hallazgos.

Aclaración de los objetivos

El primer paso es tener bien claros los objetivos del estudio por encuesta. Por ejemplo, digamos que un sistema escolar pequeño pide a un investigador estudiar las actitudes hacia el uso de castigos en las escuelas públicas. Como parte de su plan, el investigador necesita considerar la naturaleza de la pregunta que se está haciendo. ¿Lo que preocupa es la efectividad del castigo? ¿La forma como se administra el castigo? ¿El tipo de castigo (físico o de otro tipo)?

La definición de la naturaleza de los objetivos podría requerir entrevistas preliminares de personas que podrían entrevistarse más a fondo en una fase posterior del proyecto. Una de las metas primarias en este paso del proyecto es definir las variables, como castigo y actitudes, que se van a estudiar. Ambos términos, que son bastante vagos de por sí, necesitan mayor aclaración y definición para que las preguntas que finalmente haga el investigador produzcan información útil.

Identificación de la muestra

Una vez que se han especificado los objetivos, el siguiente paso es la definición de un plan de muestreo y la obtención de la muestra de individuos que participarán en el estudio. ¿Se incluirá a todos los padres y maestros? Probablemente no, porque sería una muestra demasiado grande y no resultaría eficiente encuestar a un grupo tan grande. Entonces, ¿cómo representar objetivamente a esa comunidad?

¿Regrese al capítulo 4! ¿Qué tal si tomamos una muestra aleatoria estratificada de tres padres de cada año escolar de cuatro escuelas del distrito? ¿Y una muestra aleatoria de administradores de cada uno de dos niveles administrativos, administración del edificio y administración central? Si van a participar niños, tal vez el investigador quiera diseñar un plan que tenga en cuenta la frecuencia con que han sido castigados los niños y por qué razón. Tener sólo niños que casi nunca son castigados o sólo niños que reciben frecuentes castigos introduciría un sesgo en las características de la muestra y en los resultados.

Definición del método

Ahora que los objetivos y el plan de muestreo están claros, ¿exactamente qué va a suceder durante la entrevista o panel de estudio? He aquí algunas de las cuestiones que podrían preocupar a un investigador.

¿Las preguntas serán primordialmente abiertas o cerradas, o una combinación de ambas? ¿Cómo muestreará cada pregunta el contenido, las opiniones o las actitudes?

¿Cómo se definirá la muestra de encuestados? ¿Incluirá padres o maestros o administradores o las tres cosas? ¿Y los estudiantes?

¿Cómo se recabarán los datos? ¿Se utilizarán entrevistas? ¿Encuestas por correo?
¿Qué tipos de preguntas se harán? ¿Qué información objetiva se incluirá?

Estas preguntas se contestarán, en parte, con los tipos de información que el investigador necesite para cumplir con los objetivos que se definieron en una etapa temprana del proyecto.

Codificación y calificación

Las investigaciones por encuesta pueden producir desde respuestas largas que deben analizarse hasta contestaciones sencillas de *si* o *no*, dependiendo del formato y el contenido de la pregunta. Una vez recabados los datos, el investigador necesitará codificarlos (1 = hombre, 2 = mujer; por ejemplo) y luego calificar las respuestas según un plan organizado que facilite la tabulación.

En la figura 9.1 se muestra un ejemplo sencillo. Aquí vemos un desglose de los padres que utilizan con regularidad el castigo físico y los que no lo hacen, y los juicios de cada grupo en cuanto a la efectividad del castigo físico.

Se puede realizar algún tipo de análisis de las frecuencias de estas respuestas para contestar la pregunta acerca de las actitudes de los padres hacia el castigo.

El castigo físico es cruel e ineficaz	El Castigo físico es duro e innecesario	El castigo físico puede funcionar en ciertas condiciones	El castigo físico es útil para frenar la mala conducta	El castigo físico es el método más eficaz para manejar la mala conducta
12	14	15	23	32
46	13	14	7	6

Figura 9.1 Número de padres que reaccionaron a afirmaciones acerca del castigo físico.

La validez de los datos de encuestas

La recopilación de datos de encuestas es una tarea ardua, que implica buscar constantemente sujetos y tratar con muchas fuentes ajenas de varianza difíciles de controlar. No obstante, es un tanto sorprendente lo relativamente fácil que es establecer la validez de tales datos. Por ejemplo, si entrevistamos a una persona, una forma de establecer la validez de los datos es buscar una fuente alternativa que los confirme. Es fácil cotejar los expedientes públicos para confirmar datos como la edad y la afiliación a algún partido político. Incluso se puede entrevistar otra vez a los encuestados para confirmar la veracidad de lo que dijeron la primera vez. Naturalmente, la gente puede mentir dos veces, pero un buen investigador es consciente de esa posibilidad y trata de confirmar la información objetiva que podría ser importante para los fines del estudio.

Evaluación de las investigaciones por encuesta

Al igual que todos los métodos de investigación, la investigación por encuesta tiene sus ventajas y desventajas.

En el lado positivo, este tipo de investigación permite obtener una imagen muy amplia de lo que se está estudiando. Si se muestrea correctamente, no es difícil generalizar a millones de personas, como se hace con regularidad en las encuestas realizadas durante las campañas políticas y cosas así. Esta capacidad para generalizar implica un gran ahorro en dinero y tiempo.

Segunda, la investigación por encuesta es eficiente porque la parte de recopilación de datos del estudio termina después de un contacto con los encuestados. Además, se requiere un mínimo de recursos. En algunos casos, basta con un portapapeles y un cuestionario para recabar los datos.

Tercera, si se lleva a cabo correctamente y con un error de muestreo mínimo, las encuestas pueden producir resultados notablemente exactos.

Las desventajas pueden ser serias. Las más importantes son las fuentes de predisposición, de las que hay dos tipos y que pueden surgir durante las entrevistas o en los cuestionarios. La **predisposición del entrevistador** ocurre cuando el entrevistador predispone sutilmente al encuestado para que responda de una manera u otra. Esta predisposición podría presentarse, por ejemplo, si el entrevistador alienta (aunque sea de la forma más inconsciente) la aprobación o desaprobación de una respuesta mediante una sonrisa, un fruncimiento del ceño, el desvío de la mirada o alguna otra acción. Por otra parte, la persona que está siendo entrevistada podría responder con una predisposición por no querer dar otra cosa que no sean *respuestas socialmente aceptables*. Después de todo, ¿cuántas personas responderían con un "¡sí!" definitivo a la pregunta "¿Golpea usted a su cónyuge?" Hay que cuidarse de estas amenazas de predisposición capacitando cuidadosamente a los entrevistadores para que sean objetivos, asegurándose de que las preguntas no guíen al entrevistado ni lo coloquen en una posición en la que tenga pocas opciones.

Otro problema con las investigaciones por encuesta es la posibilidad de que la gente no conteste, como en el caso de las encuestas por correo. ¿Por qué es esto importante? Porque las personas que no contestaron podrían constituir un grupo cualitativamente distinto del de aquellas que sí lo hicieron. Por tanto, los hallazgos basados en los que sí respondieron serán diferentes de los que se obtendrían si se hubiera considerado todo el grupo.

Investigación del desarrollo

El campo del psicólogo del desarrollo (y de muchos educadores, pediatras, antropólogos y demás) es entender los cambios que ocurren durante todo el proceso de desarrollo, desde la concepción hasta la muerte. Para lograr esto, dos métodos de investigación básicos han evolucionado durante los últimos 100 años para describir los cambios o diferencias en el comportamiento dentro de un marco de referencia de diferentes edades o etapas de la vida. Examinaremos cada uno de estos tipos, analizando un ejemplo, y luego hablaremos de sus ventajas y desventajas relativas.

El método longitudinal

El método longitudinal determina los cambios en el comportamiento de un grupo de sujetos en más de un punto en el tiempo. En otras palabras, si sometemos a prueba a un grupo de sujetos de 30 años de edad en 1960, y luego aplicamos la prueba al mismo grupo otra vez en 1965 (cuando tenían 35 años), otra vez en 1970 (cuando tenían 40 años) y así sucesivamente (como se ilustra en la figura 9.2),

Año de la prueba

	1960	1965	1970	1975	1980
1940	20	25	30	35	40
1935	25	30	35	40	45
Año de nacimiento (cohorte)					
1930	30	35	40	45	50
1925	35	40	45	50	<u>55</u>
1920	40	45	50	55	<u>60</u>

La edad aparece en cursivas

Corte transversal

Longitudinal

Figura 9.2 Aquí puede verse un diseño tanto longitudinal como de corte transversal en el que se da el año en que nacieron los sujetos, cuándo se realizó la prueba y su edad. La línea continua indica un estudio de corte transversal, y la punteada, un estudio longitudinal.

estaríamos realizando un estudio longitudinal. La línea punteada de la figura 9.2 indica el diseño para un estudio longitudinal en el que el mismo grupo de participantes que nacieron en 1930 se prueba cinco veces sucesivas a intervalos de cinco años.

Los estudios longitudinales se realizan para examinar los *cambios con la edad* durante un periodo prolongado. Por ejemplo, Singer y otros (1984) efectuaron un estudio longitudinal sobre televisión, imaginación y agresión. El propósito del estudio fue examinar las costumbres de ver televisión dentro de un entorno familiar y las posibles influencias que tal costumbre podría tener sobre los patrones de interacción social de la familia.

Los investigadores aplicaron pruebas a diferentes olas (o grupos) de niños, entre 1977 y 1982, contando con un grupo final de 84 niños al final del experimento. Se pidió a los padres que llevaran una bitácora diaria de los programas de televisión que vejan sus hijos y los investigadores entrevistaron a los padres, analizaron informes escolares y medidas de inteligencia, entrevistaron a los niños y obtuvieron otra información.

El método longitudinal tiene claras ventajas. La más importante es que permite estudiar el desarrollo durante un periodo prolongado. Es más, dado que las mismas personas se estudian en más de un punto en el tiempo, los sujetos actúan como sus propios controles. En otras palabras, cada persona contribuye con los mismos antecedentes (genéticos, étnicos o de otro tipo) y experiencias a la situación de prueba. Este tipo de diseño es muy potente porque se minimiza la variación intraindividual.

También existen desventajas importantes. Primera, estos tipos de estudios son muy costosos. No sólo cuesta seguir la pista a las personas durante periodos largos, sino que los costos del personal y de otro tipo se incrementan de un año a otro. Ésta es una razón por la que actualmente se inician relativamente pocos proyectos longitudinales, en comparación con los tiempos en que se iniciaron algunos de los estudios clásicos, hace muchos años. Uno de éstos, el estudio Terman de niños superdotados, se inició en la década de 1920 y todavía no termina.

Otra desventaja de los diseños longitudinales es que las personas abandonan los experimentos (fenómeno conocido como mortalidad). Aunque podríamos suponer que esta tasa de abandonos es aleatoria, a menudo hay razones para creer que los abandonos son sistemáticos. Esto implica que cierto tipo de personas podría abandonar el estudio y dejar la muestra restante con características y cualidades sustancialmente distintas de las de la muestra original.

El método de sección transversal

Mientras que el método longitudinal examina un grupo de personas repetidamente al paso del tiempo, el **método de sección transversal** examina varios grupos de personas en un solo punto del tiempo. En otras palabras (como se aprecia en la figura 9.2), si examináramos las *diferencias por edad* entre sujetos de 30, 35, 40, 45 y 50 años de edad (todos nacidos en años distintos) en 1970, estaríamos realizando un estudio de sección transversal. Este método sirve para examinar diferencias por edad, más que cambios con la edad, como se hace con el método longitudinal.

Por ejemplo, para averiguar a quién piden diferentes tipos de consejos los niños de distintas edades cuando enfrentan diversos tipos de problemas, Wintre y otros (1988) utilizaron el método de sección transversal. Ellos estudiaron 24 hombres y 24 mujeres de 8, 11, 14 y 17 años de edad, y les presentaron tres problemas hipotéticos. Los investigadores pidieron a los jóvenes escoger un familiar adulto, un adulto experto, un familiar de su misma edad o un experto de su misma edad como consultor. Aquí, los investigadores estaban examinando las diferencias por edad (no los cambios), y al seleccionar grupos con diferentes edades y evaluar sus respuestas en el mismo punto del tiempo, alcanzaron su objetivo.

Al igual que el método longitudinal, el de sección transversal tiene sus ventajas y desventajas.

El método de sección transversal es mucho más económico que el longitudinal, ya que las pruebas se realizan durante un tiempo limitado. Gracias a la brevedad de este periodo, los abandonos son mínimos. La gente tiende a permanecer en el mismo sitio durante el tiempo suficiente para terminar el proyecto. ¿Y las desventajas? La más grave es la falta de comparabilidad de los grupos, pues lo único en lo que difieren es en la edad. Y, como veremos en un momento más, la edad no es una variable independiente muy útil.

En la tabla 9.2 se muestra una comparación de las ventajas y desventajas de las estrategias de investigación longitudinal y de sección transversal.

Empleo de estudios de seguimiento

La información de la tabla 9.2 proporciona una muy buena idea de los beneficios y deficiencias de cada uno de estos métodos. La decisión respecto a cuál se usará depende de factores como la cantidad de recursos con que se cuenta, las restricciones de tiempo y, desde luego, la pregunta que se está haciendo.

	Ventajas	Desventajas
Corte transversal	Bajo costo Corta duración Baja tasa de abandonos No requiere administración ni cooperación a largo plazo entre el personal y los sujetos	Falta de comparabilidad de grupos No da idea de la dirección de cambio que un grupo podría tomar Examina personas con la misma edad cronológico que podrían tener diferentes edades de maduración Nada dice acerca de la continuidad del desarrollo en casos individuales.
longitudinal	Abundantes detalles sobre el proceso de desarrollo Alta comparabilidad de los grupos Permite estudiar la continuidad entre edades muy distintas Permite hacer especulaciones de causa y efecto modificadas acerca de la relación entre variables	Costoso Posible alta tasa de abandonos

Tabla 9.2 Comparación de los diseños de investigación de corte transversal y longitudinal, con sus ventajas y desventajas.

Casi nunca resulta práctico realizar cualquier tipo de estudio longitudinal como parte de la experiencia escolar en el nivel de licenciatura o posgrado, ya que el tiempo requerido para recopilar los datos suele ser demasiado largo. En cambio, los estudios de seguimiento son muy factibles, pues en ellos se pueden aprovechar datos ya recopilados como base para recopilar datos adicionales.

Por ejemplo, veamos un estudio clásico de 25 bebés criados en un orfanatorio en el que recibieron cuidados básicos aceptables pero muy poca atención y afecto humanos. Harold Skeels (1942) hizo que transfirieran 13 de esos bebés del orfanatorio a una institución para mujeres con retraso mental, donde fueron "adoptados" por las mujeres. Skeels observó que los niños que fueron criados en la institución para mujeres y recibieron estímulos obtuvieron en promedio 28 puntos más en las pruebas de IQ que los niños que se dejaron en el orfanatorio. Skeels llevó a cabo la parte de seguimiento del estudio 21 años después (y él mismo publicó los resultados del seguimiento en 1966), cuando trató de determinar si los diferentes cuidados que los grupos de niños habían recibido habían tenido algún efecto duradero. Para su deleite, encontró que todos los bebés que habían formado parte del grupo experimental (los que habían sido transferidos) eran autosuficientes, 11 de ellos se habían casado y nueve de éstos tenían hijos. Tristemente, un tercio de los niños que no recibieron experiencias especiales todavía seguían en instituciones como adultos, y sólo unos cuantos de los niños que habían permanecido en el orfanatorio estaban llevando una vida adulta normal. Skeels no siguió a estos sujetos durante toda su vida, pero sí realizó un estudio de seguimiento que proporcionó información de naturaleza longitudinal.

El papel de la edad en el estudio del desarrollo

La edad es una variable curiosa, y la gente se apoya mucho en ella para explicar los cambios observados en una amplia variedad de conductas humanas. Por ejemplo, aunque podría ser cómodo describir los cambios en la forma como los niños usan las palabras en las diferentes edades en función de la edad, probablemente sería más correcto entender dichos cambios en términos de los cambios en la complejidad cognoscitiva, experiencia y la capacidad para manipular símbolos.

En otras palabras, la edad tiene un valor descriptivo pero no necesariamente un valor explicativo. Aunque la edad puede describir lo que está sucediendo, por sí sola no puede decirnos por qué. Don Baer (1970) resume esta observación en un artículo muy persuasivo, "An Age Irrelevant Concept of Development". Él argumenta que la experiencia, no la edad, es la fuerza impulsora de las diferencias observadas en el desarrollo, y que estudiar dichas diferencias es mucho más fructífero que estudiar el comportamiento en función de la edad cronológica o de maduración.

Estas observaciones y preocupaciones generales acerca de la utilidad de la edad han dado pie a tipos adicionales de diseños para investigación del desarrollo además de los métodos longitudinal y de sección transversal que acabamos de describir. Algunas de estas técnicas nuevas toman en cuenta variables tales como cuando se midió la conducta (los llamados efectos del tiempo de medición) y los efectos de la cohorte (o del grupo). Tomemos un estudio básico de sección transversal que examina grupos de personas nacidas en diferentes años y a los que se les aplicó la prueba en la misma fecha. Si encontramos diferencias entre los grupos (o diferencias de edad), ¿cómo sabemos que las distinciones no se deben al año en que nacieron, más que a la edad? ¿Cómo podría la fecha de nacimiento contribuir a tales diferencias? Muy fácilmente. ¿Qué tal si un grupo de personas nació antes de que se descubriera un fármaco o una técnica o incluso antes de que ocurriera cierto suceso cultural que facilita o dificulta el aprendizaje? Tomemos, por ejemplo, los niños que nacieron antes y después de "Plaza Sésamo", el programa intensivo de enriquecimiento cognoscitivo preparatorio transmitido por televisión. Bien podría ser que ver ese programa haya tenido un impacto significativo sobre las habilidades de lenguaje. En este caso, la cohorte (año de nacimiento) y la edad podrían confundirse. Hay confusión cuando dos variables (como la fecha de nacimiento y la edad) explican la misma cosa (diferencias en las habilidades de lenguaje) y no es posible separar los efectos de las dos.

Otro ejemplo de confusión se presenta con la edad y la fecha en que se realizó la medición en un estudio llevado a cabo por Nesselroade y Baltes (1974). Ellos examinaron los cambios en la personalidad durante la adolescencia y encontraron disminuciones con la edad en medidas de la fortaleza del superego, ansiedad y aprovechamiento. Aunque podría ser tentador atribuir esos cambios a la edad, los científicos también encontraron que, sea cual fuera la edad del niño durante el periodo de examen de tres años (algunos pasaron de los 13 a los 15 años mientras que otros pasaron de los 16 a los 18 años), la caída en los puntajes era la misma. Obviamente, el cambio de edad no era importante. Lo que sí fue importante fue el "momento cultural" en el que se determinaron las conductas. Éste es un ejemplo de influencia histórica. Lo que estaba sucediendo en el momento de realizarse la prueba al parecer afectó los puntajes de los niños independientemente de su edad.

Aunque los estudios de desarrollo que utilizan la edad de maduración o cronológica como variable dependiente importante pueden *describir* de manera aceptable los cambios con el tiempo, hay que tener cuidado de que otros factores, como los señalados en el estudio de Nesselroade y Baltes, no sean también atractivos como fuentes de explicación.

Investigación correlacional

La **investigación correlacional** describe la relación lineal entre dos o más variables sin tratar en lo absoluto de atribuir a una variable el efecto observado en otra. Como técnica descriptiva, la investigación correlacional es muy potente porque indica si dos variables (como el número de horas de estudio y el puntaje obtenido en una prueba) tienen algo en común. Si así es, existe una correlación entre las dos.

Tal vez recuerde que en el capítulo 5 utilizamos el coeficiente de correlación para estimar la confiabilidad de una prueba. Aquí se usa la misma medida, también en un sentido descriptivo. Por ejemplo, se utilizan correlaciones como medida estándar para determinar la

relación entre el grado de parentesco (gemelos, primos, no parientes, etc.) y la similitud de los puntajes en pruebas de inteligencia. Cuanto mayor es la correlación, más alto es el grado de parentesco. En tal caso, esperaríamos que gemelos criados en el mismo hogar tuvieran puntajes de IQ más parecidos (tienen más en común) que gemelos criados en diferentes hogares, ¡y así es! Los gemelos que se crían separados sólo comparten la misma herencia genética, mientras que los criados en el mismo hogar comparten antecedentes tanto hereditarios como ambientales.

La relación entre variables

La medida más frecuente de las relaciones es el coeficiente de correlación. Un **coeficiente de correlación** es un índice numérico que refleja la relación entre dos variables; se expresa como un número entre -1.00 y +1.00 y aumenta su valor a medida que aumenta la cantidad de varianza que una variable comparte con otra. Es decir; cuanto más tienen en común dos cosas (como los gemelos idénticos), más fuertemente tienden a estar relacionados entre ellas (lo cual es lógico). Si usted comparte intereses con otra persona, es más probable que sus actividades estén relacionadas, que si se compara con alguien con quien usted nada tiene en común.

Por ejemplo, es más probable encontrar una relación más fuerte entre los puntajes obtenidos en una prueba de destreza manual y en una de coordinación ojo-mano que entre los puntajes de una prueba de destreza manual y la altura de la persona. Asimismo, cabe esperar que la correlación entre los puntajes obtenidos en pruebas de matemáticas y de lectura sea más fuerte que entre los puntajes obtenidos en pruebas de lectura y de fuerza física. Esto se debe a que los desempeños en las pruebas de matemáticas y de lectura tienen más en común (habilidades intelectuales y de solución de problemas, por ejemplo) que en una prueba de lectura y una de levantamiento de pesas.

Las correlaciones pueden ser *directas* o *positivas*, lo que significa que, a medida que una variable cambia de valor; la otra cambia en la misma dirección, como la relación entre el número de horas que usted estudia y su calificación en un examen. Cuanto más estudia usted, más alta es la calificación que obtiene. Asimismo, cuanto menos estudia usted, más baja es su calificación. Observe que la palabra *positiva* a veces se interpreta como sinónimo de *buena*. Éste no es el caso aquí.

Las correlaciones también pueden reflejar una relación *indirecta* o *negativa*, lo que significa que, a medida que una variable cambia de valor en una dirección, la otra cambia en la dirección opuesta, como la relación entre la rapidez con que usted contesta reactivos de opción múltiple y el puntaje que obtiene en la prueba. Cuanto más rápidamente trabaje (en general), más bajo será su puntaje. Si contesta con más detenimiento, es probable que su puntaje sea mayor. No interprete esto como que si usted trabaja lentamente será más inteligente. Las cosas no funcionan así, lo que es un ejemplo adicional de cómo las correlaciones no son causales. Lo que significa es que para un grupo dado de estudiantes, la correlación entre el tiempo de presentación de una prueba y el puntaje total es negativa. Puesto que se trata de una estadística de grupo, es difícil concluir algo acerca del desempeño de un individuo en particular y es imposible atribuir causalidad. Hablaremos más de esto después.

Resulta interesante que la cualidad importante de un coeficiente de correlación no es su signo, sino su valor absoluto. Una correlación de -.78 es más fuerte que una de +.68, así como una correlación de +.56 es más débil que una de -.60.

Aspecto de los coeficientes de correlación

La medida de relación de uso más frecuente es la correlación de momento producto de Pearson, que se representa con la letra r minúscula seguida de símbolos que representan las variables que se están correlacionando. El símbolo r_{xy} representa una correlación entre la variable x y la variable y. Para poder calcular una correlación es preciso tener un par de puntajes (como un puntaje en lectura y uno en matemáticas) para cada sujeto del grupo con el que se está trabajando.

Como acabamos de mencionar, las correlaciones pueden variar entre -1.00 y +1.00, y pueden adoptar cualquier valor entre esos dos extremos. Por ejemplo, examine la figura 9.3, donde se muestran cuatro conjuntos de datos (A, B, C y D), cada uno representado por un diagrama de dispersión. Un diagrama de dispersión es una gráfica de los puntajes en pares. En el conjunto A, la correlación es de +.70. (Veremos cómo se calcula ese valor en un momento más).

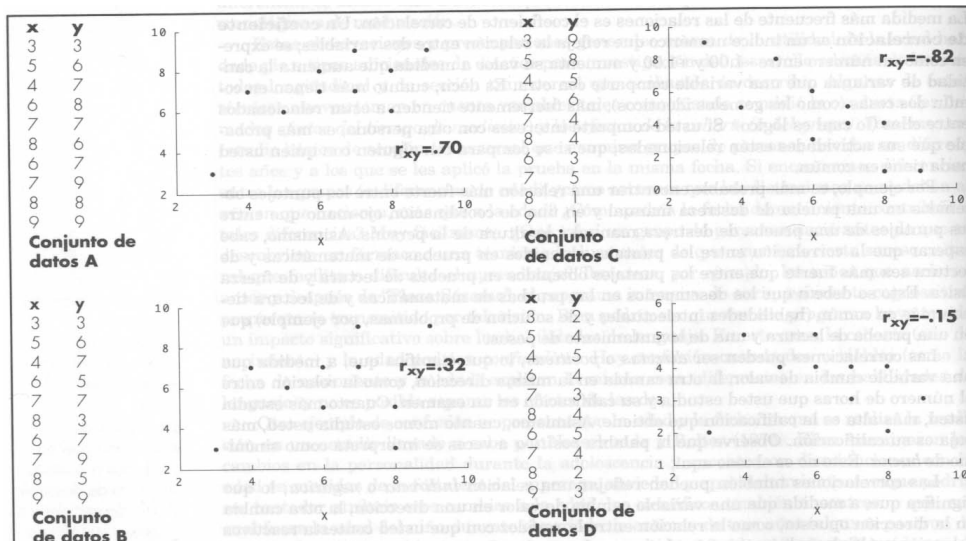


Figura 9.3 diferentes diagramas de dispersión y correlaciones

Para dibujar un diagrama de dispersión, siga estos pasos:

1. En papel milimétrico, trace un par de ejes x (horizontal) y y (vertical).
2. Indique cuál variable del par será x y cuál será y.
3. Para el Sujeto 1, marque las coordenadas de los valores x y y. En este ejemplo (conjunto de datos A de la figura 9.2), el puntaje x es 3 y el puntaje y es 3, así que se marcó un dato puntual correspondiente a (3,3).
4. Repita el paso 3 para todos los datos puntuales y verá el diagrama de dispersión que se muestra en la figura 9.2 para el conjunto de datos A.

Ahora examine el conjunto de datos B, en el que la correlación es de sólo .32, mucho más débil que la de .70. Podemos ver que la correlación más fuerte (conjunto A) se caracteriza por lo siguiente:

- Los datos puntuales se agrupan más cerca de una línea recta imaginaria a medida que aumenta la fuerza de la correlación.
- A medida que la inclinación de dicha línea imaginaria se acerca a un ángulo de 45°, la correlación se hace más fuerte.

Los datos del conjunto A exhiben una correlación positiva elevada, mientras que los del B muestran una correlación positiva más baja. Los datos del conjunto C muestran una correlación negativa elevada (-.82) y, tal como sucede con una correlación positiva elevada, las coordenadas que representan la intersección de dos datos puntuales se alinean a lo largo de una diagonal (en este caso de la esquina superior izquierda a la inferior derecha, acercándose a un ángulo de 45°) El último conjunto de datos, el D, muestra muy poca relación ($r_{xy} = -.15$) entre las variables x y y, y la gráfica correspondiente de las coordenadas revela un patrón débil. En otras palabras, una línea trazada por estos puntos sería casi plana u horizontal.

En síntesis, cuanto más marcada es la formación de un patrón y más se alinea dicho patrón en un ángulo de 45° (sea desde la esquina inferior izquierda de la gráfica a la esquina superior derecha, en el caso de las correlaciones positivas, o desde la esquina superior izquierda a la esquina inferior derecha, en el caso de las correlaciones negativas), más claros serán los indicios visuales de la existencia de una relación entre dos variables.

Cálculo de coeficiente de correlación de Pearson

La forma manual más fácil de calcular la correlación entre dos variables es utilizando el método del puntaje bruto. La fórmula para r_{xy} (donde el subíndice xy representa la correlación entre x y y) es

$$r_{xy} = \frac{\sum (X-X)(Y-Y)}{\sqrt{\sum (X-X)^2 \sum (Y-Y)^2}}$$

donde: r_{xy} = el coeficiente de correlación
 X = el puntaje de un individuo en la variable x
 Y = el puntaje de un individuo en la variable y

\bar{X} = media de la muestra para la variable x
 \bar{Y} = media de la muestra para la variable y

Veamos un ejemplo sencillo en el que se calcula el coeficiente de correlación a partir del conjunto de datos C que se muestra en la figura 9.2. La media de la variable x es 6.3 y la de la variable y es 4.6. He aquí la ecuación final:

$$R_{XY} = \frac{-37}{\sqrt{(32.1)(62.4)}}$$

Pruébelo usted mismo y vea si puede obtener la misma respuesta ($r_{xy} = -.82$)

La correlación es la expresión de la relación entre las variables x y y, y se indica con r_{xy} . ¿Qué sucede si tenemos más de dos variables? En tal caso tendremos más de un coeficiente de correlación. En general, si tenemos n variables, tendremos " n tomadas de dos en dos" pares de relaciones. En la tabla 9.3 se muestra una matriz de correlación, o una tabla que revela las correlaciones por pares entre tres variables (año escolar, puntaje en lectura y puntaje en matemáticas). Cada uno de los tres coeficientes de correlación se calculó empleando la misma fórmula que describimos antes.

	Año escolar	Lectura	Matemáticas
Año escolar	1.000		
Lectura	0.321	1.000	
Matemáticas	-0.039	0.605	1.000

Tabla 9.3 Las correlaciones de año escolar, puntaje en lectura y puntaje en matemáticas.

Quizá haya notado usted que la diagonal de la matriz se llena con valores 1.000 porque la correlación de cualquier cosa consigo misma siempre es 1. Además, los coeficientes a la derecha de la diagonal y a su izquierda forman una imagen especular. Es decir, si llenamos las correlaciones para la otra "mitad" de la matriz (arriba de la diagonal de valores 1.000, en la tabla 9.3), se repetirán los mismos valores.

Interpretación del coeficiente de correlación

El coeficiente de correlación es un índice interesante: refleja el grado de relación entre variables, pero es relativamente difícil de interpretar sin más información. No obstante, hay dos formas en que podemos interpretar estos indicadores de relaciones generales.

La primera es el método "a simple vista", en el que correlaciones con ciertos valores se asocian a cierto grado nominal de relación, de modo que

<i>Las correlaciones entre....</i>	<i>se consideran</i>
<i>.8 y 1.0</i>	<i>muy fuertes</i>
<i>.6 y .8</i>	<i>fuertes</i>
<i>.4 y .6</i>	<i>moderadas</i>
<i>.2 y .4</i>	<i>débiles</i>
<i>.0 y .2</i>	<i>muy débiles</i>

Recuerde: no se deje engañar por estos números. Incluso la correlación más débil (digamos .1) puede ser estadísticamente significativa si la muestra en que se basa es lo bastante grande y se acerca lo suficiente al tamaño de la población. Ya explicamos la distinción entre significatividad y significado en el capítulo anterior.

Un método más sólido para interpretar el coeficiente de correlación es elevar al cuadrado su valor y calcular el **coeficiente de determinación**. Este valor, r_{xy}^2 es la cantidad de varianza de una variable que se explica por la otra. En otras palabras, este coeficiente nos permite estimar la cantidad de varianza que se puede explicar en una variable examinando la cantidad de varianza de otra variable. Así, pues, si la correlación entre dos variables es .40, el coeficiente de determinación es .16.16% de la varianza de una variable se puede explicar por la varianza de la otra variable. Por lógica, 84% de la varianza no se ha explicado. Esta porción de varianza no explicada se conoce como **coeficiente de alienación**.

Una comparación interesante es ver cómo cambia la cantidad de varianza explicada en la relación entre dos variables a medida que la correlación se hace más fuerte.

En la tabla 9.4 podemos ver el coeficiente de correlación simple (columna A) y el coeficiente de determinación (columna B). Estudie el cambio en la cantidad de varianza explicada a medida que aumenta el valor del coeficiente de correlación. Por ejemplo, si la correlación aumenta de .4 a .5, el aumento en la cantidad de varianza explicada es de 9%. Pero si la correlación se aumenta una cantidad similar

(digamos, de .6 a .7), el aumento en la cantidad de varianza explicada es de 13%. El aumento en la varianza explicada no es lineal. Cuanto más alta es la correlación, mayor es el "salto" en la varianza explicada.

r_{xy}	r_{xy}^2	Cambio en la varianza explicada
.1	1%	
.2	4%	3%
.3	9%	5%
.4	16%	7%
.5	25%	9%
.6	36%	11%
.7	49%	13%
.8	64%	15%
.9	81%	17%
1.0	100%	19%

Tabla 9.4 Diferencias en la cantidad de varianza explicada en función de diferentes valores del coeficiente de correlación. La cantidad de varianza explicada aumenta al aumentar la correlación.

¿Lo que usted necesita es un diseño experimental? Ésta no es realmente la pregunta que debe hacerse. Más bien, ¿su tema de interés exige usar las herramientas sugeridas por el método histórico o descriptivo? Como subrayamos antes, la pregunta que se hace determina la forma en que se contesta. Si usted quiere investigar cómo los colonos de Oklahoma en la década de los treinta criaban a sus niños, o cómo ha cambiado la crianza de los niños, ¿la historiografía podría ser para usted! ¿Y qué ofrece el método descriptivo? Ofrece un relato de un suceso, a menudo con tanto detalle que sirve como punto de partida para plantear y contestar otras preguntas. Los estudios de casos, la investigación del desarrollo y los estudios correlacionales describen un fenómeno específico de tal manera que comunica la imagen global de lo que se está estudiando. Aunque estos métodos no permiten darse el lujo de implicar alguna relación de causa y efecto entre variables, su uso proporciona las herramientas necesarias para responder preguntas que de otro modo no podrían contestarse.

Ejercicios

- ¿Cuál es el objetivo primario de la investigación histórica?
- Mencione algunas de las diferencias más importantes que los investigadores históricos usan, en comparación con los investigadores más tradicionales (experimentales) en las ciencias sociales y de la conducta.
- Describa, en un párrafo, un estudio de investigación histórica que a usted le gustaría llevar a cabo. Conteste las preguntas siguientes:
 - ¿Cómo establecería la autenticidad de sus fuentes?
 - ¿Cómo establecería la exactitud de sus fuentes?
- Escriba una hipótesis nula y una de investigación para un estudio longitudinal y uno de sección transversal. Dé ejemplos de estudios en los que usted querría examinar los cambios con la edad o las diferencias por edad.
- Ordene los coeficientes de correlación siguientes del más fuerte al más débil.
 - .21
 - .67
 - .53
 - .01
 - .78
- ¿Qué defecto tiene el argumento siguiente? El número de horas que usted estudia tiene una relación directa con lo bien que a usted le va en las pruebas escolares. Por tanto, si usted tiene un mal resultado en un examen, quiere decir que no estudió lo suficiente.
- Mencione algunos de los problemas que implica el uso de la edad como variable dependiente e indique cómo podrían resolverse.
- Indique el tipo de correlación que describe cada una de las relaciones siguientes: positiva, negativa o ninguna relación.
 - Al aumentar A, el valor de B aumenta.
 - Al aumentar A, el valor de B disminuye.
 - Al disminuir el valor de A, B no cambia.
 - Al disminuir A, el valor de B aumenta.

9. Indique si las hipótesis siguientes son de naturaleza correlacional.
 - a. No hay diferencias en la capacidad cognoscitiva entre los niños en edad preescolar que asisten a guarderías privadas y los niños en edad preescolar que no asisten a guarderías.
 - b. Existe una relación entre la educación de los padres, su situación socioeconómica y el aprovechamiento de sus hijos en matemáticas.
 - c. No hay relación entre la tasa de crímenes violentos en Nueva York y la situación socioeconómica.
 - d. La educación de los padres no mejora el desempeño de un niño en una prueba de aprovechamiento de matemáticas.
 - e. A lo largo del tiempo, hay diferencias en las políticas de disciplina empleadas en las escuelas rurales y urbanas.

10. ¿Cuál es el propósito de la investigación descriptiva?

11. ¿Qué diferencia hay entre el método de sección transversal y el método longitudinal? Identifique dos ventajas y dos desventajas de cada método.

12. ¿Cuáles de las afirmaciones siguientes acerca de los coeficientes de correlación son verdad?
 - a. Las correlaciones pueden ser positivas.
 - b. Las correlaciones pueden ser negativas.
 - c. Las correlaciones reflejan causas.
 - d. Las correlaciones miden la relación entre dos variables.

CAPITULO DIEZ

Métodos de investigación experimental

En Neil J. Salkind “Métodos de Investigación”, Prentice Hall, México, 1998.

Lo que aprenderá en este capítulo

- La importancia y el papel de los diseños experimentales
- La importancia de la aleatoriedad en el método experimental
- El papel del azar en el método experimental
- Los principios del diseño experimental
- Los conceptos de validez interna y externa, y el papel que desempeñan en el método experimental
- Las amenazas a la validez interna y externa, y cómo se pueden controlar
- Cómo controlar las fuentes extrañas de variabilidad

Lo que los científicos normalmente tratan de hacer es averiguar por qué suceden las cosas. Ellos hacen hasta lo imposible por establecer, por ejemplo, por qué algunos niños son más activos que otros, por qué algunos adultos tienen más éxito, o a qué se deben las diferencias de actitud. Los métodos y modelos que describiremos en este capítulo pueden ayudar mucho a entender tales fenómenos.

Una herramienta que puede ser útil para entender tales diferencias es el **método experimental**. A diferencia de los demás métodos que hemos descrito hasta ahora en este libro, el método experimental trata de determinar la presencia de una *causa y un efecto definidos*. Esto implica que una vez que se usa este método puede emitirse un juicio acerca de que si A causa que B suceda o A no causa que B suceda. Otros métodos, como los modelos histórico y descriptivo, no nos ofrecen ese lujo. Aunque esos métodos pueden servir para descubrir relaciones entre variables, no hay forma de establecer una relación causal. ¿Por qué? Es gracias al método experimental, que nos permite controlar las posibles fuentes de las diferencias (o la varianza), que podemos afirmar lo siguiente: un factor está relacionado con otro de tal manera que los cambios en ese factor tienen una relación causal con los cambios en el otro.

Por ejemplo, el diseño experimental más sencillo sería uno en que dos grupos de sujetos se escogen aleatoriamente de una población y un grupo (llamado **grupo experimental**) recibe un tratamiento, mientras que el otro grupo (llamado **grupo control**) no lo recibe. Al final del experimento, se aplica la prueba a ambos grupos para ver si hubo diferencia en sus puntajes. Suponiendo (y éste es el gran supuesto) que los dos grupos eran equivalentes en el momento en que se inició el experimento, cualquier diferencia observada al final del experimento debe ser consecuencia del tratamiento. De esto se trata el diseño experimental, que puede adoptar varias formas.

Si se ponen en práctica correctamente, los diseños experimentales pueden proporcionar un altísimo nivel de potencia y control que nos permite entender las relaciones causales entre variables. Su uso es, en buena medida, responsable de gran parte de los conocimientos que los científicos han adquirido acerca de los procesos psicológicos y sociales.

Diseños experimentales

Hay diversos tipos de diseños experimentales. En esta sección describiremos el esquema que Donald Campbell y Julian Stanley hicieron famoso en su monografía publicada en 1963 (*Experimental and Quasi-Experimental Design for Research on Teaching*) que ayudó a revolucionar la forma en que la gente planea y lleva a cabo sus proyectos de investigación. Ellos identificaron tres categorías generales de

diseños de investigación: preexperimental, experimental verdadero y cuasiexperimental. Los diseños cuasiexperimentales también se conocen como diseños causales-comparativos. En este capítulo estudiaremos los diseños preexperimentales y experimentales verdaderos, y en el siguiente trataremos el diseño cuasiexperimental.

La principal diferencia entre estos tipos de diseños experimentales radica en el grado de control que imponen a las variables que se están estudiando. El método preexperimental tiene el grado más bajo de control, el verdadero método experimental tiene el grado más alto y el método cuasiexperimental está en un punto intermedio. Cuanto mayor es el control, más fácil es atribuir una relación de causa y efecto a una secuencia de sucesos.

Otro aspecto en el que estos diseños preexperimental, experimental verdadero y cuasiexperimental difieren entre ellos es el grado de aleatoriedad que entra en el diseño. Ya sabemos que la palabra *aleatorio* implica una probabilidad igual e independiente de ser seleccionado, pero esa definición y concepto se pueden aplicar más allá de la selección de sujetos de una población para formar una muestra, a la importancia del concepto en el diseño experimental.

De hecho, hay varias medidas que deben tomarse para asegurar la cualidad de verdadera aleatoriedad en los mejores diseños experimentales.

La primera es la que más conocemos: la selección aleatoria de sujetos de una población para formar una muestra. Este es el primer procedimiento que realizaríamos en un experimento. Ahora ya tenemos una muestra.

Segunda, es importante asignar aleatoriamente los sujetos a los diferentes grupos. Por ejemplo, debemos asegurarnos de que los sujetos asignados al grupo 1 tuvieron la misma probabilidad de ser asignados al grupo 2.

Por último (si hicimos lo que recomiendan los pasos 1 y 2), tenemos dos grupos que suponemos que son prácticamente equivalentes entre ellos. Ahora necesitamos decidir cuál de los dos grupos recibirá el tratamiento o, si tenemos cinco grupos, cuál tratamiento recibirá cada grupo. De la misma forma como usamos antes una tabla de números aleatorios, simplemente asignamos (al azar) diferentes tratamientos a los grupos.

Después de seguir estos tres pasos podemos estar seguros de que los sujetos se seleccionaron aleatoriamente de una población y que se asignaron al azar a grupos, y que también se decidió en forma aleatoria cuál grupo recibía cuál tratamiento.

En la tabla 10.1 se resumen algunas de las diferencias primordiales entre los diseños preexperimental, experimental verdadero y cuasiexperimental. Aunque dejaremos la explicación de los diseños cuasiexperimentales para el siguiente capítulo, los incluimos aquí para que usted pueda comparar todos los tipos de diseños. Observe que muchas de las diferencias tienen que ver con el proceso de aleatoriedad de los procedimientos de selección, los sujetos y la asignación.

	Diseños preexperimentales	Diseño cuasiexperimental	Diseños experimentales verdaderos
¿Presencia de un grupo control?	En algunos casos, pero normalmente no	Sí	Sí
¿Selección aleatoria de sujetos de una población?	No	Sí, pero restringida en los grupos preasignados	Sí
¿Asignación aleatoria de sujetos a grupos?	No	No	Sí
¿Asignación aleatoria de tratamientos a grupos?	No	Sí es posible	Sí
Grado de control sobre las variables extrañas	Bajo	Moderado	Alto

Tabla 10.1 Diferencias entre diseños preexperimentales, experimentales verdaderos y cuasiexperimentales.

Diseños preexperimentales

Los **diseños preexperimentales** no se caracterizan por la selección aleatoria de participantes de una población, ni incluyen un grupo control. Sin estas dos cosas, la capacidad de investigación para descubrir la naturaleza causal de la relación entre las variables independiente y dependiente se reduce considerablemente, y casi se elimina. Estos diseños tienen poco o ningún control sobre las variables extrañas que podrían ser responsables por resultados distintos de lo que pretendía el investigador.

Por ejemplo, un padre utiliza un antiguo remedio familiar (¡llevar un collar de ajos!) para alejar los malos espíritus asociados al resfriado de uno de sus hijos. ¡Y he aquí que funciona! Éste es el tipo de conclusión más débil al que podemos llegar ya que no se efectúa comparación alguna para demostrar que el ajo funcionó mejor que cualquier otra cosa o mejor que no hacer nada. El niño, por supuesto, podría haber mejorado por sí solo. Sencillamente no hay control sobre otros factores que podrían causar el resultado observado (como que el virus del resfriado haya seguido su curso natural).

En términos de investigación, este tipo de estudio se denomina **diseño de estudio de caso de un solo intento** y tiene este aspecto:

Grupo 1	Tratamiento	Prueba posterior
---------	-------------	------------------

Como puede verse, se somete a un grupo a algún tipo de tratamiento y después se le aplica la prueba.

¿Qué deficiencias percibe usted en este tipo de diseño preexperimental de caso de estudio de un solo intento? Primera, en ningún momento se trata de imprimir un carácter aleatorio. ¿De qué podría servir un diseño preexperimental como este estudio de caso de un solo intento? No de mucho para realizar trabajos experimentales o establecer relaciones de causa y efecto, pero es aceptable si estamos especulando acerca de factores que ocurrieron en una ocasión anterior y su efecto en una conducta posterior.

Otro diseño preexperimental, llamado **diseño con prueba previa y posterior de un grupo** se representa así:

Grupo 1	Prueba previa	Tratamiento	Prueba posterior
---------	---------------	-------------	------------------

Por ejemplo, a un investigador le interesa estudiar qué tan efectivo es el método A para aumentar la fuerza muscular. Él o ella sigue estos pasos para realizar el experimento:

1. Solicita voluntarios para el experimento con un anuncio.
2. Administra una prueba previa para medir la fuerza.
3. Expone los sujetos al tratamiento que según la hipótesis aumentará la fuerza.
4. Administra la prueba posterior.

Las comparaciones importantes son entre los puntajes de la prueba previa y la posterior para cada sujeto. El problema primordial con este tipo de diseño es que no hay grupo control, y sin dicho grupo ¿cómo puede el investigador saber si cualquier diferencia observada entre los puntajes de la prueba previa y la posterior es función del tratamiento o de otros factores? ¿Qué tal si 50% de la muestra no durmió lo suficiente antes de la prueba posterior? ¿O si ellos participaron en otro estudio diseñado también para aumentar la fuerza? Estos factores, más que el tratamiento específico, podrían ser los responsables de las diferencias en la fuerza física.

Diseños experimentales verdaderos

Los verdaderos diseños experimentales incluyen todos los pasos de selección y asignación de sujetos de manera aleatoria, más un grupo control, con lo que proporcionan un argumento más fuerte para postular una relación de causa y efecto. Una de las razones por las que estos diseños son tan potentes es que todos realizan la selección y asignación de tratamientos y a grupos de manera aleatoria.

Por ejemplo, examinemos uno de los más populares de estos diseños, el diseño con **prueba previa y posterior y grupo control**, que tiene este aspecto:

Grupo 1	Prueba previa	Tratamiento	Prueba posterior
Grupo 2	Prueba previa	Ningún tratamiento	Prueba posterior

Los pasos que se siguen cuando se emplea este diseño son:

1. Asignar aleatoriamente los sujetos al grupo experimental o al de control.
2. Realizar la prueba previa de la variable dependiente con cada grupo.
3. Aplicar el tratamiento al grupo experimental. El grupo control no recibe el tratamiento.
4. Realizar la prueba posterior de la variable dependiente tanto con el grupo experimental como con el de control (en otra forma o formato si es necesario).

El supuesto aquí, y es probable que usted ya lo haya captado, es que dado que los sujetos se asignan aleatoriamente al grupo control o bien al experimental, ambos grupos son equivalentes al principio del experimento. Cualesquier diferencias que se observen al final del experimento se deben al tratamiento porque todas las demás explicaciones ya se descartaron.

Los diseños de prueba previa y posterior con grupo control no están limitados tampoco a dos grupos. Por ejemplo, digamos que un investigador desea examinar los efectos de diferentes programas de alfabetización sobre lo bien que los adultos aprenden a leer. Un tratamiento podría ser dar instrucción cinco días a la semana y otro podría ser darla tres días a la semana. El tercer grupo sería un grupo control, cuyos miembros no reciben instrucción alguna.

El diseño experimental se vería más o menos así:

Grupo 1	Prueba previa	5 días/semana	Prueba posterior
Grupo 2	Prueba previa	3 días/semana	Prueba posterior
Grupo 3	Prueba previa	Ningún tratamiento	Prueba posterior

El número de grupos de tratamiento no afecta realmente en tanto exista un grupo control. Existe una diferencia importante en cuanto a la naturaleza del grupo control. En algunos casos, el grupo control podría no tener tratamiento alguno, mientras que en otros podría recibir un tipo diferente de tratamiento. La diferencia en el papel del grupo control es un reflejo del tipo de pregunta que se hizo originalmente.

Si el grupo control no recibe ningún tratamiento, la pregunta obvia es si el tratamiento es eficaz en comparación con la ausencia total de tratamiento. Si el grupo de tratamiento se compara con otro grupo que también recibe tratamiento, la pregunta pasa a ser ¿cuál de los dos es más eficaz? Es una distinción un tanto sutil, pero es importante recordarla al pensar en cómo estructurar el experimento.

He aquí otro verdadero diseño experimental que es muy popular: el diseño Solo **con prueba posterior y grupo control**, cuyo aspecto es el siguiente:

Grupo 1	Tratamiento	Prueba posterior
Grupo 2	Ningún tratamiento	Prueba posterior

La característica más evidente aquí es que no se administra la prueba previa ni al grupo control ni al experimental. La justificación de este enfoque es que si se seleccionan las personas y se asignan a grupos de forma aleatoria, no hay necesidad de efectuar una prueba previa, pues de todos modos los grupos ya son equivalentes, ¿o no? La respuesta es "sí" cuando se tiene una muestra suficientemente grande (al menos 30 en cada grupo). Otra razón para usar el diseño sólo con prueba posterior en lugar del diseño con prueba previa y posterior es que a veces no es conveniente o siquiera posible administrar una prueba previa. En tales condiciones, puede optarse por el diseño sólo con prueba posterior.

No obstante, el uso de un diseño sólo con prueba posterior tiene dos desventajas básicas. Primera, si los procedimientos de aleatoriedad no fueron apropiados, los grupos podrían no ser equivalentes por principio. Segunda, no se puede usar la prueba previa para asignar las personas a otros grupos experimentales, digamos alto o bajo según alguna variable. Estas desventajas podrían no ser importantes, pero siempre deben considerarse.

El último diseño experimental verdadero es algo así como el abuelito de todos los demás: el **diseño de Solomon de cuatro grupos**, que mostramos aquí:

Grupo 1 (experimental)	Prueba previa	Tratamiento	Prueba posterior
Grupo 2 (de control)	Prueba previa	Ningún tratamiento	Prueba posterior
Grupo 3 (de control)	Sin prueba previa	Tratamiento	Prueba posterior
Grupo 4 (de control)	Sin prueba previa	Ningún tratamiento	Prueba posterior

Este diseño incluye cuatro grupos, uno experimental (que recibe el tratamiento) y tres grupos control, uno de los cuales también recibe el tratamiento.

Lo más interesante y útil de este diseño son los muchos tipos de comparaciones que pueden efectuarse para determinar qué factores podrían ser los causantes de ciertos tipos de resultados. Por ejemplo, digamos que a usted le interesa determinar los efectos del tratamiento pero también quiere saber si el hecho mismo de someterse a una prueba previa altera los puntajes finales. Para determinar la efectividad del tratamiento podríamos comparar el grupo experimental con el grupo 2, ya que la única diferencia entre los dos es la inclusión de un tratamiento. Si se desea determinar la influencia de la prueba previa sobre resultados posteriores, una comparación del grupo 1 y el grupo 3 proporcionaría la información necesaria. La única diferencia es que un grupo realizó la prueba previa, mientras que el grupo 3 no la hizo. Hay muchas otras comparaciones que pueden hacerse. Por ejemplo, una comparación de los grupos 3 y 4 daría información sobre el

efecto del tratamiento sobre grupos que no presentaron la prueba previa (pero si recibieron el tratamiento).

Entonces, ¿por qué no todos los que efectúan investigación experimental verdadera usan este tipo de diseño? Una razón es el tiempo. Si bien el diseño experimental de Solomon de cuatro grupos es muy bueno para separar los factores que causan las diferencias en la variable dependiente, su ejecución consume mucho tiempo. Hay que organizar cuatro grupos seleccionados y asignados aleatoriamente, tres condiciones de control y una experimental, y realizar muchas pruebas. Hoy día no abunda el tiempo ni el dinero, así que para muchos investigadores esta clase de diseño no es práctica.

Validez interna y externa y diseño experimental

Los diferentes tipos de diseños experimentales que hemos mencionado en este capítulo se bosquejaron en los trabajos fundamentales de Campbell y Stanley. Ellos se dieron cuenta de que no bastaba con idear diferentes diseños, sino que también se necesitaba una forma de evaluar dichos diseños. ¿Qué criterios externos podríamos usar para juzgar la utilidad de las diferentes maneras de atacar un problema?

¿Su decisión? Utilizar los criterios de validez interna y externa; ambos miden qué tan bien el diseño hace lo que debería hacer.

La **validez interna** es la cualidad de un diseño experimental que hace que los resultados obtenidos se deban a la manipulación de la variable independiente. Por ejemplo, si podemos demostrar que un tratamiento logra incrementar las habilidades sociales de niños retraídos, y si el tratamiento es la única causa aparente del cambio, decimos que el diseño (y el experimento) es internamente válido. Si hay varias explicaciones distintas para los resultados de un experimento, éste no tiene validez interna.

La **validez externa** es la cualidad de un diseño experimental que permite generalizar los resultados, de la muestra original a otra muestra y luego por extensión a la población de la cual se obtuvo la muestra. Por ejemplo, si usted puede aplicar el tratamiento para aumentar las habilidades sociales de los niños retraídos a otro grupo de niños retraídos, decimos que el diseño (y el experimento) tiene validez externa.

Sin embargo, no todos los diseños y no todos los experimentos tienen niveles aceptables de validez interna y externa por muy diversas razones, que Campbell y Stanley llaman *amenazas para la validez interna y externa*. Una vez que entendamos estas amenazas, podremos determinar cuáles diseños experimentales son preferibles y por qué.

Amenazas para la validez interna

La que sigue es una breve explicación de las amenazas para la validez interna que reducen la probabilidad de que los resultados del experimento se deban a la manipulación de la variable independiente. Los buenos científicos tratan de reducir estas amenazas.

Historia

Muchos experimentos se realizan durante un tiempo prolongado, y podrían ocurrir otros sucesos externos al experimento que podrían afectar sus resultados. Dichos sucesos podrían ofrecer una explicación más atractiva de las diferencias observadas entre los grupos, que el tratamiento original.

Por ejemplo, un investigador quiere estudiar el efecto de dos dietas distintas sobre la conducta de niños hiperactivos en la escuela. Sin que el investigador lo sepa, algunos de los padres de los niños del grupo experimental se han puesto en contacto con el maestro de los niños, y juntos han iniciado un programa en el hogar para reducir las conductas problemáticas en la escuela. Si hubo una diferencia en cuanto a la conducta escolar de los niños que participaron en el plan de dieta, ¿cómo sabemos que no se debió a la colaboración entre el maestro y los padres? Esa influencia exterior (la actividad del maestro y los padres) es un ejemplo de la **historia** como amenaza para la validez interna, ya que la actividad fuera de la escuela, y no la dieta, podría explicar cualquier diferencia observada.

Maduración

La **maduración** puede definirse como los cambios causados por fuerzas biológicas o psicológicas. Estos cambios podrían opacar los que son el resultado de un tratamiento.

Por ejemplo, una investigadora está estudiando los efectos de un programa de entrenamiento, que dura un año, sobre la fuerza física de niños en edad escolar. Al término del programa ella evalúa la fuerza de los niños y encuentra que el puntaje de fuerza promedio aumentó en el plazo de un año. ¿Su conclusión? Que el programa funcionó. ¿Correcto? Tal vez. Pero por atractiva que sea esa explicación, la naturaleza misma del desarrollo físico hace que la fuerza de los niños aumente con la edad o maduración.

Selección

La base de cualquier experimento es la **selección** de los sujetos como participantes. La selección es una amenaza para la validez interna de un experimento cuando el proceso de selección no es aleatorio, sino que contiene una predisposición sistemática que podría hacer que los grupos participantes sean diferentes unos de otros.

Por ejemplo, un investigador quiere determinar cómo el cuidado después de la escuela afecta la cohesión familiar. Como parte del experimento, el investigador forma un grupo experimental (aquellas familias cuyos niños reciben cuidado prolongado) y uno de control

(aquellas familias cuyos niños no reciben dicho cuidado). Puesto que las familias no se seleccionaron ni se asignaron a los tratamientos de manera aleatoria, no hay forma de saber si eran equivalentes entre ellas. El grupo de niños que recibió cuidado prolongado podría provenir de familias con una actitud positiva o negativa hacia el programa, aun antes de que éste iniciara, lo que predispondría los resultados.

Pruebas

En muchos experimentos psicológicos, una prueba previa forma parte del experimento. Cuando la prueba previa afecta el desempeño en mediciones posteriores (como en una prueba posterior), las **pruebas** pueden ser una amenaza para la validez interna.

Por ejemplo, un investigador aplica una prueba previa a un grupo de sujetos para medir sus habilidades en matemáticas de segundo de secundaria, y luego les enseña (el tratamiento) una nueva forma de resolver ecuaciones elementales. Se administra la prueba posterior, y hay un incremento en la proporción de respuestas correctas. Dada esta información, no sabemos si el aumento se debe al aprendizaje de una nueva forma de resolver las ecuaciones elementales o al aprendizaje que podría haber ocurrido durante la presentación de la prueba previa. La experiencia con la prueba previa podría haber bastado para hacerlos "conocer la prueba" y lo que su desempeño posterior refleja es eso, más que la efectividad del tratamiento.

Instrumentación

Cuando se afecta la forma misma de obtener un puntaje con un instrumento, cualquier cambio en los puntajes podría deberse al procedimiento de calificación más que a los efectos del tratamiento.

Por ejemplo, un investigador está utilizando una prueba de redacción para juzgar la efectividad de un programa para mejorar la escritura. Es casi seguro que para cuando califique el centésimo examen estará usando un conjunto de criterios distinto del que usó cuando calificó el primero. Incluso si los criterios no cambian, la simple fatiga seguramente afectará el juicio del calificador y producirá diferencias debidas a la instrumentación y no a los efectos reales del programa.

Regresión

Ésta es una amenaza en verdad fascinante (que a veces no se entiende bien). El mundo de la probabilidad está construido de tal manera que la ubicación en cualquiera de los extremos de un *continuum* (como un puntaje muy alto o muy bajo) tiene como resultado puntajes que acusan una regresión hacia la media en una prueba subsecuente (empleando la misma prueba). En otras palabras, cuando los niños obtienen un puntaje muy alto o muy bajo en alguna medida, cabe esperar que sus puntajes subsecuentes se desplazarán hacia la media, más que en la dirección opuesta. Esto sólo se cumple si su ubicación original (en el extremo) fue resultado de su puntaje en la prueba. El lector quizá se haya dado cuenta ya de que la regresión ocurre a causa de la falta de confiabilidad de la prueba y del error de medición que se introduce y que coloca a las personas más hacia los extremos de lo que probablemente es el sitio que les pertenece.

Dada la menor probabilidad de que alguien quede en la parte extrema de una distribución (sea baja o alta), es más probable que, en pruebas adicionales, obtenga un puntaje situado en un área más central respecto a la distribución. En el caso de sujetos con puntajes altos o bajos, moverse hacia el centro de la distribución implica moverse hacia la media; esto es de lo que se trata la regresión.

Por ejemplo, una maestra de niños con discapacidades físicas severas diseña un proyecto dirigido a mejorar su capacidad para cuidarse ellos mismos y somete el grupo a una prueba previa empleando información anecdótica, en septiembre, antes de que el programa principie. La maestra sólo incluye en el proyecto aquellos niños que obtuvieron puntajes muy bajos en la prueba. En junio, ella aplica la prueba otra vez a los sujetos y observa que sus habilidades han aumentado. Un argumento sólido que podría presentarse es que el aumento se debió a **regresión** y no a algo que la maestra haya hecho. Es decir, los niños que inicialmente estaban en un extremo (en cuanto a la prueba de capacidad para cuidarse ellos mismos) se desplazarían hacia el puntaje medio (volviéndose menos extremos) aunque nada suceda. El cambio ocurre únicamente por regresión y podría no tener nada que ver con el tratamiento.

Mortalidad

Uno de los problemas del mundo real que afectan a la investigación es que a veces es difícil localizar los sujetos para realizar estudios de seguimiento. Cuando esto sucede, el investigador debe preguntarse si la composición del grupo después del abandono de algunos miembros sigue siendo básicamente la misma que era antes. La **mortalidad** es una amenaza para la validez interna de un experimento cuando los miembros que abandonan alteran la naturaleza del grupo mismo.

Por ejemplo, la investigación con bebés es fascinante pero a menudo frustrante. Los bebés casi siempre llegan dormidos o llorando o listos para comer pero rara vez listos para «jugar», y muchos tienen que enviarse de vuelta a casa y reprogramarse o hasta excluirse del experimento. Los bebés que desertan podrían ser muy diferentes de los que se quedan, y la muestra final de sujetos deja de ser equivalente a la inicial. Esto pone en entredicho la efectividad del tratamiento sobre esta muestra distinta. La mortalidad también se denomina *bajas*.

Amenazas para la validez externa

Así como hay amenazas para la validez interna de un diseño, existen también para la validez externa. Una vez más, la validez externa no se ocupa de si la manipulación de la variable independiente tuvo o no algún efecto sobre la variable dependiente (eso está dentro del ámbito de la validez interna), sino de si los resultados de un experimento pueden o no generalizarse a otro entorno. He aquí las amenazas para la validez externa, junto con definiciones y ejemplos. Al igual que con las amenazas para la validez interna, los buenos científicos tratan de reducir el peligro que representan.

Interferencia de múltiples tratamientos

Un conjunto de sujetos podría recibir un tratamiento "inadvertido" además del tratamiento intencional, lo que reduce la generalizabilidad de los resultados a otro entorno en el que el tratamiento inadvertido tal vez no esté disponible.

Por ejemplo, digamos que un grupo de residentes de un asilo de ancianos está aprendiendo a ser más energético, y uno de los asistentes se interesa por el programa y "enseña" algo a los residentes por su cuenta. Los resultados del experimento no podrían generalizarse fácilmente a los residentes de otro asilo, ya que en ese otro asilo podría o no haber algún asistente igualmente emprendedor.

Disposiciones reactivas

En la planta manufacturera Hawthorne de Chicago, Illinois, ciertos investigadores se propusieron hace unos 50 años medir los efectos de modificar ciertos indicios ambientales, iluminación y horas de trabajo, sobre la producción de los trabajadores. El problema fue que los participantes del estudio conocían las intenciones de los investigadores. Incluso cuando la iluminación era más pobre y la jornada de trabajo era más larga, la productividad del grupo experimental aumentó. ¿Por qué? Porque los obreros recibieron atención especial de los investigadores. Dicha atención causó cambios en la productividad, y se determinó que la iluminación y las condiciones de trabajo tenían una importancia secundaria. Como los sujetos no se estudiaron en otros entornos (lo que anularía el propósito del experimento), la validez externa fue baja, lo mismo que la generalizabilidad.

Por cierto, esta amenaza para la validez externa, las disposiciones reactivas, también se conoce como (¡adivinó usted!) efecto Hawthorne.

Efectos de experimentador

Otra amenaza para la validez externa tiene que ver con los investigadores. Imagine un experimento encaminado a reducir la ansiedad que produce la necesidad de acudir al dentista. (¡Qué horror!) ¿Qué tal si la persona que está impartiendo la capacitación desensibiliza-dora se estremece cada vez que el dentista pone en marcha su taladro? Los resultados de semejante programa de capacitación no podrían generalizarse a otro entorno, pues para ello sería necesario que en ese otro entorno hubiera un instructor que se comportara de forma similar. De lo contrario, la naturaleza de la experiencia habrá cambiado. En otras palabras, el programa de capacitación podría no ser tan efectivo sin las expresiones emocionales del instructor, y los resultados del programa de capacitación podrían no ser generalizables ya que la persona que imparte la capacitación no forma parte del programa.

Sensibilización por prueba previa

Ya hemos visto cómo las pruebas previas pueden informar a las personas de lo que está por venir y afectar sus puntajes futuros, lo que reduce la validez interna de un estudio. De forma similar, la presencia de una prueba previa puede alterar la naturaleza del tratamiento, de modo que si el tratamiento se aplica en otro entorno sea menos o más efectivo a causa de la ausencia de la prueba previa. Para hacer equivalentes las situaciones y maximizar la generalizabilidad a otros entornos, la prueba previa tendría que ser parte del tratamiento, lo que, por definición, alteraría la naturaleza del tratamiento y el propósito del experimento.

Como aumentar la validez interna y externa

Primero, la validez interna. No es secreto cómo se puede maximizar la validez interna de un experimento: seleccione al azar sujetos de una población, asígnelos aleatoriamente a grupos y utilice un grupo control. En casi todos los diseños en los que están presentes las características anteriores, se eliminan casi todas las amenazas. Tomemos el ejemplo de los niños con discapacidades físicas severas y el proyecto que inicia en septiembre con el fin de aumentar la capacidad de los niños para cuidar de sí mismos. Si se incluye un grupo que no recibe el programa (el grupo control), el supuesto es que tanto el grupo control como el

Diseño ↓	Amenaza →												
	Historia	Maduración	Selección	Pruebas	Instrumen-tación	Regre-sión	Mortali-dad	Interferencia de múltiples tratamientos	Disposi-ciones reactivas	Efectos de experimen-tador	Sensibiliza-ción por prueba previa		
Diseños preexperimentales	Estudio de caso de un solo intento	-	-	-									
	Diseño con prueba previa y posterior de un grupo	-	-	+	-	-	?	+					
Verdaderos diseños experimentales	Diseño con prueba previa y posterior y grupo control	+	+	+	+	+	+	+	?	?			
	Diseño sólo con prueba posterior y grupo control	+	+	+	+	+	+	+	?	?		+	
	Diseño de Solomon de cuatro grupos	+	+	+	+	+	+	+	?	?		+	

Notas:
+ La amenaza se toma en cuenta.
? Dependiendo del entorno y las circunstancias, las amenazas podrían haberse tomado en cuenta.
- Las amenazas no se toman en cuenta.
Un espacio en blanco indica que la amenaza no es relevante.

experimental exhibirán el mismo avance o regresión, de modo que cualquier diferencia que se observe al final del año deberá ser resultado del programa de capacitación.

Asimismo, si los grupos son equivalentes en un principio (lo que se asegura con la aleatoriedad), los cambios se deberán al tratamiento y no a la falta de equivalencia al inicio del experimento.

La inclusión de un grupo control y el empleo de aleatoriedad también da cuenta de las otras amenazas, como las pruebas, la mortalidad y la maduración. Suponiendo que los grupos son inicialmente iguales y que se exponen a circunstancias y experiencias similares, las únicas diferencias entre ellos serían en función del tratamiento, ¿no?

Asegurar la validez externa es un asunto un tanto diferente porque está menos ligada al diseño que al comportamiento de las personas que realizan el experimento. Por ejemplo, la única forma de asegurar que los efectos experimentales no sean una amenaza para la validez externa del experimento es asegurarse de que el investigador que administra el tratamiento actúe de tal manera que no lo interfiera. En el ejemplo relativo a la desensibilización de pacientes dentales nerviosos, hay que asegurarse de que el instructor no tenga problemas con el entorno del consultorio dental.

Si bien la mayor parte de las amenazas para la validez interna se elimina cuidando el diseño del experimento, el diseñador mismo debe ocuparse de eliminar las amenazas a la validez externa.

En la tabla 10.2 se resumen las amenazas para la validez interna y externa de los diseños que tratamos en este capítulo, y se indica qué tan bien cada diseño compensa cada amenaza.

Validez interna o externa: ¿trueque?

Si tenemos cuidado, ésta podría ser una situación en la que no sea necesario sacrificar una cosa para obtener la otra.

Un experimento puede ser válido tanto interna como externamente, pero sólo si tenemos cuidado y balanceamos los factores. Cuanto mayor es la validez interna, más confianza podemos tener en que lo que hicimos (manipular la variable independiente) causó los resultados que observamos. Sin embargo, si el control es excesivo (como cuando se usan procedimientos experimentales muy estrictos con una muestra de sujetos definida de manera muy específica) los resultados del experimento podrían ser difíciles de generalizar (y por tanto tener menos validez externa) a cualquier otro entorno. La razón es que el grado de control tan alto podría ser imposible de duplicar, para no hablar de lo difícil que sería encontrar una muestra similar a la que se usó originalmente.

¿La solución? Utilice su criterio. Procure realizar sus experimentos de tal manera que aseguren un grado moderado de validez interna mediante el control de las fuentes extrañas de varianza usando aleatoriedad y un grupo control. Lo mismo podemos decir de la validez externa. Si no podemos generalizar a otros grupos, el valor de nuestra investigación (dependiendo de su propósito) podría ser limitado.

Como controlar las variables extrañas

Hemos hablado una y otra vez de variables extrañas, pero ¿qué son? Las variables extrañas son factores que pueden reducir la validez interna de un estudio. Se trata de variables que, si no se toman en cuenta de una forma u otra, pueden confundir los resultados. Como vimos en el capítulo 9, los resultados se confunden cuando no es posible separar los efectos que diferentes factores podrían tener sobre algún resultado. Por ejemplo, un investigador está estudiando los efectos de los desayunos escolares sobre la asistencia a la escuela. Los padres que están más motivados podrían preocuparse más por que sus niños vayan a la escuela y reciban los desayunos, y esto podría explicar la diferencia entre los que asisten y los que no lo hacen. Podría ser que el desayuno por sí mismo nada tenga que ver con cualquier diferencia entre grupos. En este caso, el tratamiento (el desayuno) se confunde con la motivación de los padres.

En casi cualquier investigación experimental que examinemos, encontraremos variables que potencialmente podrían confundir los resultados. Estas variables enturbian las aguas cuando un científico trata de entender qué factores exactamente causan qué resultados. ¿Cómo puede resolverse este problema? Hay varias soluciones. La pregunta general debe ser: *¿Por cuáles variables hay que preocuparse y cuáles se puede suponer que carecen de importancia?* En el caso de las variables que sí preocupan, ¿qué puede hacerse para minimizar el efecto que podrían tener sobre los resultados del experimento?

En primer lugar; podemos decidir hacer caso omiso de cualquier variable que no esté relacionada con la variable dependiente que se está midiendo. Por ejemplo, si la asistencia es la variable dependiente primaria y el ofrecimiento de un almuerzo escolar es la variable independiente primaria, ¿podrían ser importantes factores tales como el género del niño, el género del maestro, el tamaño del grupo o la edad de los padres? Podría ser. La única forma de comprobarlo en un sentido o en otro es haciendo una revisión bibliográfica e ideando algún argumento conceptual sólido que permita afirmar que el género del maestro, digamos, está o no relacionado con la asistencia del niño. En general, si no puede argumentarse de manera convincente que una variable está relacionada con el resultado que se está estudiando, es probable que lo mejor sea no tomarla en cuenta.

En segundo lugar la aleatoriedad permite controlar los efectos de muchas fuentes potenciales de varianza. Lo más importante es que la aleatoriedad ayuda a asegurar que los vez grupos experimentales y de control sean equivalentes en cuanto a muy diversas características. En el ejemplo anterior; la asignación aleatoria de los niños a los grupos que reciben y que no reciben desayunos aseguraría que la motivación de los padres tenga la misma probabilidad de influir en ambos grupos y, por tanto, no sea muy atractiva como explicación de cualquier diferencia observada.

Igualación

La asignación aleatoria de sujetos a grupos es una buena forma de asegurar que, en general, los grupos sean equivalentes. No obstante, podría haber ocasiones en las que el investigador quiera asegurarse de que los dos grupos sean iguales en lo que respecta a un atributo, rasgo o característica específico. Por ejemplo, en el estudio del programa de desayunos escolares, si la influencia de los padres es causa de preocupación y el investigador no cree que la asignación aleatoria resuelva el problema, podría usar la técnica de igualación.

La igualación de sujetos implica sencillamente que, por cada ocurrencia de un individuo con un puntaje de X en el grupo experimental, el investigador se asegurará de que haya una persona en el grupo control con un puntaje similar. En general, la regla que debemos recordar es que la variable respecto a la cual los sujetos se van a igualar debe estar fuertemente relacionada con la variable dependiente de interés; de lo contrario, la igualación no tiene mucho sentido. Dado que ésta es la regla general, no debe sorprender mucho que el primer paso del proceso de igualación sea obtener una medida de la variable que se igualará antes de efectuar la asignación a grupos. Luego se ordenan los puntajes y se seleccionan pares de sujetos que quedan juntos. Un sujeto de cada par se coloca en cada grupo, y el experimento continúa.

Lo que están haciendo los investigadores cuando siguen esta estrategia es acomodar los naipes en su favor para asegurar que ciertas influencias importantes y posiblemente potentes no afecten desmedidamente los resultados del estudio. Ésta es una forma sencilla y efectiva de lograrlo.

Sin embargo, como seguramente sospechará usted, la igualación tiene su lado negativo. Esta técnica puede ser costosa y consumir mucho tiempo, y cabe la posibilidad de no encontrar una pareja para todos y cada uno de los individuos. Supongamos que un conjunto de padres está extremadamente motivado y el conjunto con la siguiente motivación más alta quedó muy por debajo del primero en la escala. ¿Podemos igualar esos conjuntos? Probablemente no. Tal vez lo que tenga que hacerse es excluir a los padres que tuvieron el puntaje extremo o encontrar otro conjunto que obtenga un puntaje tan alto como ellos para poder igualarlo con el otro.

Empleo de grupos homogéneos

Una de las mejores maneras de asegurar que las variables extrañas no sean un factor importante es utilizar una población homogénea, o una cuyos miembros sean muy parecidos, y seleccionar la muestra de ella. De este modo, casi todas las fuentes de diferencias (como antecedentes raciales o étnicos, educación, actitud política, etc.) podrían quedar controlados automáticamente. Una vez más, es muy importante que los grupos sean homogéneos sólo en aquellos factores que podrían afectar sus puntajes de la variable dependiente.

ANCOVA

Una última técnica es un procedimiento relativamente avanzado llamado análisis de covarianza, o ANCOVA. Esta herramienta estadística nivela cualesquier diferencias iniciales que pudieran existir. Por ejemplo, digamos que usted está investigando si un programa de ejercicios especializado aumenta la velocidad al correr. Puesto que usted sabe que la velocidad tiene cierta relación con la fuerza, desea asegurarse de que los participantes en el programa tengan todos más o menos la misma fuerza. Digamos que usted trata de igualar sujetos pero descubre que la diversidad es demasiado grande para asegurar que la igualación nivelará los grupos. Entonces, usted decide olvidarse de la igualación y utilizar un análisis de covarianza, o ANCOVA.

El análisis de covarianza, en su nivel más básico, resta al efecto del tratamiento la influencia de la relación entre la covariable (que en este caso es la fuerza) y la variable dependiente (que aquí es la velocidad). En otras palabras, ANCOVA ajusta los puntajes de velocidad finales de modo que reflejen la posición inicial de los participantes en cuanto a la fuerza. ANCOVA es una técnica especialmente útil en los diseños cuasiexperimentales o causales-comparativos en los que no es fácil asignar aleatoriamente los sujetos a los grupos pero se tiene información acerca de las variables que están relacionadas con el resultado final y en las cuales los sujetos difieren.

¿Quiere averiguar si A causa B? Los métodos experimentales son justo lo que el doctor le recetó. Estos métodos proporcionan un grado de control que es difícil alcanzar empleando cualquiera de los otros métodos que habíamos estudiado hasta ahora en el libro. Los trabajos decisivos de Campbell y Stanley identificaron las diversas amenazas para estos diseños y proporcionaron herramientas para evaluar la validez interna y externa de diversos diseños preexperimentales y experimentales. Gracias a técnicas como la igualación, el uso de grupos homogéneos y algunas técnicas estadísticas, podemos tener mucha confianza en que la diferencia entre los grupos es resultado de la manipulación de la variable independiente y no de alguna otra fuente de diferencias. Si de lo que se trata es de establecer una relación de causa y efecto, llegó usted al sitio indicado cuando llegó a este capítulo.

Ejercicios

- Defina cada una de las siguientes amenazas para la validez externa, y dé un ejemplo de cómo cada una podría ser un factor en un experimento.
 - Sensibilización por prueba previa
 - Disposiciones reactivas
- ¿Por qué es necesario lograr un balance entre la validez externa y la interna, para que una investigación sea aceptable?
- Un conjunto de niños con discapacidades emocionales se inscribe en un programa especial para mejorar la calidad de sus interacciones sociales, utilizando como criterio para inscribir a un niño en el programa el hecho de haber obtenido un puntaje extremo en cierta prueba. Al final del programa, el aumento promedio en la calidad de sus interacciones es de 57%. ¿Qué amenaza para la validez interna menoscaba el valor de este hallazgo, y qué puede hacerse para remediar la situación?
- Haga una lista con los pasos que seguiría para asegurar que dos grupos que participan en un estudio de actitud hacia el divorcio sean equivalentes.
- Escriba un resumen que describa un estudio en el que la regresión sea una amenaza para la validez interna del estudio. Asegúrese de describir las medidas que el investigador podría tomar para anular la regresión como amenaza.
- ¿Qué consideraciones éticas son importantes cuando se asignan niños de primer año de primaria a diferentes grupos de aprendizaje experimentales? ¿El hecho de que la asignación sea aleatoria reduce en algo la preocupación por los problemas de ética?
- Dé algunos ejemplos de investigación preexperimental que encuentre en el mundo real.
- Cite algunos beneficios del uso de pruebas previas.
- ¿Qué amenazas existen para la validez externa? ¿Cómo podría cada una de ellas afectarlo a usted como investigador? Dé un ejemplo de investigación para cada amenaza a la validez externa. Cite algunos métodos para aumentar la validez interna.

CAPITULO ONCE

Investigación cuasiexperimental: un pariente cercano de la investigación experimental

En Neil J. Salkind “Métodos de Investigación”, Prentice Hall, México, 1998.

Lo que aprenderá en este capítulo

- La diferencia entre los diseños experimentales y los causales-comparativos
- En qué se diferencian los diseños cuasiexperimentales de los preexperimentales y los verdaderamente experimentales

- En qué difieren unos diseños cuasiexperimentales de otros
- Empleo de diseños de sujeto único en experimentos
- Evaluación de diseños de sujeto único

En el capítulo 10 vimos que un diseño experimental puede servir para investigar la relación de causa y efecto que podría existir entre dos variables. Hay otro tipo de diseño de investigación que también trata de establecer una relación de causa y efecto pero que no tiene la fuerza ni la potencia del método experimental tradicional. En este capítulo exploraremos el método cuasiexperimental como alternativa de los diseños experimentales que ya conocemos. Aunque el método cuasiexperimental no sea tan potente, es el diseño preferido cuando en el proceso de toma de decisiones respecto a cuál diseño escoger intervienen importantes cuestiones culturales y éticas. En breve hablaremos más sobre este último punto.

El método cuasiexperimental

El **método cuasiexperimental** difiere de los métodos preexperimentales y experimentales en un aspecto muy importante.

En la investigación cuasiexperimental, el suceso que, según la hipótesis, causa las diferencias que podríamos observar entre los grupos ya ocurrió. Por ejemplo, si examinamos las diferencias entre hombres y mujeres en cuanto a habilidad verbal, la posible causa de tales diferencias (la variable independiente, el género) ya "ocurrió". En otras palabras, *ya se realizó la asignación a grupos*.

No tuvimos control sobre quién estaría en qué grupo porque el género está predeterminado, lo mismo que la edad, grupo étnico, color del cabello y cientos de otras variables. En otras palabras, hay una *preasignación* a grupos basada en algunas características del 1 grupo. Cuando usamos el método experimental (que describimos en el capítulo 10), podríamos tener una gama infinita de valores de la variable independiente para escoger. Cuando usamos el método cuasiexperimental no podemos escoger dichos valores, ni nosotros ni nadie más. Los valores de la variable independiente están ahí desde el principio, como en el caso del género (masculino y femenino), raza (caucásica, asiática, etc.), edad (menos de 18 y 18 años o más) y enfermedad (antecedentes de enfermedad cardíaca y sin enfermedad cardíaca).

Esta preasignación a grupos (o tratamientos) representa la principal deficiencia del método cuasiexperimental en comparación con el experimental clásico: es menos potente para entender la causa de cualesquier diferencias que podrían observarse en la variable dependiente. Por ejemplo, si se encuentran diferencias entre hombres y mujeres en cuanto a habilidad verbal, nuestra conclusión de que la diferencia se debe a las diferencias de género podría ser correcta, pero conceptualmente el argumento deja mucho que desear. ¿A qué puede atribuirse esta diferencia entre los sexos? ¿A la forma en que se les trata durante la juventud? ¿A las experiencias y oportunidades que tuvieron o no tuvieron? ¿A diferencias hormonales que afectan el desarrollo cerebral? Éstas son sólo tres clases de factores que podrían explicar la diferencia. Para entender plenamente la naturaleza de las diferencias habría que estudiar éstos y otros factores.

Entonces, ¿cuándo resulta útil la investigación cuasiexperimental, y cuándo debe preferirse? A pesar de las dudas que acabamos de expresar, el método cuasiexperimental es indispensable por una razón: nos permite explorar temas que de otra manera no podrían explorarse debido a cuestiones éticas, morales y prácticas. Dé un vistazo a los temas de investigación o siguientes y trate de pensar cómo podría entender sus orígenes:

- diferencias en las personalidades de niños que sufrieron abusos y que no los sufrieron efectos de la desnutrición sobre bebés
- efectos del uso materno de cocaína durante el tercer trimestre del embarazo, sobre el comportamiento del recién nacido
- diferencias en la capacidad intelectual de ancianos enviados a un asilo y ancianos que viven con su cónyuge en su propio hogar

La lista es interminable. ¿Puede ver la razón por la que se prefiere la investigación cuasiexperimental al método experimental en estos casos? Es muy obvia; todos estos ejemplos incluyen "tratamientos" o asignaciones a grupos que la ética no permitiría a un investigador aplicar artificialmente. Colocar a un niño en el grupo A (que recibe ayuda para leer mejor) o en el grupo B (que no la recibe) es una cosa, pero ¿podría usted justificar privar a una mujer embarazada de alimento para examinar los efectos sobre el bebé; o bien, enviar a un anciano a un asilo para ver los efectos que ello tendría sobre su capacidad intelectual? Jamás. Los estudios cuasiexperimentales permiten estudiar los efectos de tales variables *cuando ya han ocurrido*; es por ello que también se conocen como *investigación ex post facto* (es decir, después del hecho).

Como veremos, los diseños cuasiexperimentales contemplan la asignación aleatoria de personas a grupos, como cuando se escogen 50 de 500 hombres para formar el grupo A. Pero no es posible asignar aleatoriamente "tratamientos" a los grupos (ya están asignados), y ésta es la principal desventaja.

En términos de control y validez interna, los estudios cuasiexperimentales tienen un nivel más alto de validez interna que los diseños preexperimentales (que, como seguramente recuerda, carecen de grupo control), pero no tan alto como el de los verdaderos diseños experimentales (que tienen tanto grupo control como asignación aleatoria de tratamientos a grupos). Además, los diseños cuasiexperimentales pueden tener niveles sustanciales de validez externa, tal vez tan altos como los de los verdaderos diseños experimentales.

Diseños cuasiexperimentales

Las características más deseables de todo buen diseño de investigación son la selección y la asignación aleatorias de sujetos y el uso de un grupo control. Esto es deseable porque asegura que los grupos sean equivalentes antes de aplicar el tratamiento.

Sin embargo, hay casos en los que la aleatoriedad sencillamente no es práctica ni posible, y el uso de un grupo control es imposible o demasiado costoso, o no es razonable. Por ejemplo, no podemos decidir aleatoriamente cuáles familias que esperan un hijo van a tener varones y cuáles tendrán mujeres. Tampoco podemos decidir cuáles niños recibirán educación preescolar y cuáles no. Los diseños en los que es imposible asignar de manera aleatoria participantes a todos los grupos se denominan diseños cuasiexperimentales porque no son verdaderamente experimentales. El argumento en cuanto a las relaciones de causa y efecto en los diseños cuasiexperimentales simplemente no es tan convincente como en los verdaderos diseños experimentales.

En esta sección del capítulo veremos algunos de los diseños cuasiexperimentales más utilizados.

El Diseño de grupo control no equivalente

El **diseño de grupo control no equivalente** es uno de los diseños cuasiexperimentales que más a menudo se usan, sobre todo cuando es imposible o muy difícil asignar aleatoriamente sujetos a grupos. Por ejemplo, en un entorno educativo no sería fácil reacomodar a los niños en diferentes grupos escolares, pero nos gustaría poder usarlos como parte de una muestra. El diseño de grupo control no equivalente tiene el siguiente aspecto:

Grupo 1	Prueba previa	Tratamiento	Prueba posterior
Grupo 2	Prueba previa	Ningún tratamiento	Prueba posterior

Tal vez lo primero que usted note es lo parecido que es este diseño a aquél con prueba previa y posterior y grupo control que vimos en el capítulo 10, a excepción de que no hay selección ni asignación aleatorias. El investigador utiliza grupos intactos, como residentes de un asilo de ancianos, niños de un salón u obreros de una fábrica. Esta situación reduce de inmediato la potencia del diseño para establecer una relación causal, ya que hay dudas acerca de la equivalencia de los grupos antes de que se inicie el experimento. Es por ello que se llama diseño no equivalente.

La amenaza más seria para la validez interna de este diseño es la selección, puesto que los grupos podrían diferir inicialmente en cuanto a características que podrían estar relacionadas con la variable dependiente. Con la inclusión de una prueba previa, es posible comparar los puntajes obtenidos en dicha prueba y ver si los grupos son equivalentes. Si lo son (esto es, si no hay diferencias significativas entre ellos), no habrá que preocuparse tanto (pero sí un poco) por su equivalencia. Estadísticamente, es posible trabajar con diferencias empleando técnicas como el análisis de covarianza. Pero aunque sea posible nivelar estadísticamente las diferencias iniciales en la prueba previa, eso no significa que no existan otros factores (que la aleatoriedad podría haber anulado) que todavía podrían representar una amenaza para la validez interna del experimento.

Por ejemplo, supongamos que usted está estudiando el efecto de diferentes programas contra la discriminación racial en escuelas secundarias. Dado que no es fácil asignar niños a diferentes grupos (pues no se pueden cambiar de grupo escolar tan fácilmente), se usa el programa 1 en cierto salón y el programa 2 en otro. Estos grupos intactos no son equivalentes. Aunque se utilice algún método estadístico para hacer que sean equivalentes (como una prueba previa que evalúe la predisposición), sigue siendo poco probable que se puedan compensar otras diferencias como raza o las experiencias previas o actitudes de los niños, todos ellos factores que podrían nivelarse mejor (aunque tal vez no del todo) empleando la aleatoriedad.

El **diseño de grupo control no equivalente** es el diseño más comúnmente utilizado cuando no es posible la aleatoriedad. Funciona porque hay cierto control sobre la influencia de las variables extrañas (gracias al uso de un grupo control). Aunque no se asegura la equivalencia de los grupos, puede aproximarse en alguna medida.

La comparación con grupo estático

¿Qué tal si no es posible la aleatoriedad y tampoco administrar una prueba previa? En tal caso el diseño preferido es la **comparación con grupo estático**, que tiene el aspecto siguiente:

Grupo 1	Tratamiento	Prueba posterior
Grupo 2		Prueba posterior

Este diseño es similar al de grupo control no equivalente, excepto que no hay prueba previa. ¿En qué condiciones podríamos necesitar este diseño? Por la razón que sea, podría no haber tiempo de administrar una prueba previa, o podría ser demasiado costoso, o la muestra podría no estar disponible antes de iniciarse el tratamiento. Éstos son sólo algunos ejemplos de casos en los que el diseño de comparación con grupo estático podría ser apropiado.

¿Qué problemas tiene este diseño? Muchos. Hay poco control de las principales amenazas para la validez interna, como selección y mortalidad. En cuanto a la validez externa, todas las amenazas (interferencia de múltiples tratamientos, disposiciones de entorno reactivas

y efectos del experimentador) persisten. Por ejemplo, digamos que se está probando un tratamiento para que los residentes de asilos de ancianos aumenten su interacción social con otros residentes. Se están usando tres asilos distintos y es preciso usar el mismo tratamiento (uno de dos grupos de tratamiento y un grupo control) en cada asilo. Si se encuentra una diferencia en las habilidades sociales después del tratamiento, ¿cómo saber si ésta no se debió a diferencias que existían antes de iniciarse el experimento? No se sabe. Al no tener información de una prueba previa, no se puede determinar.

¿Cuándo querríamos usar este diseño? Cuando no hay otra opción, y ésta es una lección importante que debemos aprender acerca de estos diseños subóptimos. Se usan cuando las circunstancias impiden el uso de diseños verdaderamente experimentales, y los resultados de tales experimentos se interpretan dentro del marco de las limitaciones.

En la tabla 11.1 se presenta un resumen de las amenazas a la validez interna y externa de los diseños cuasiexperimentales y cómo se enfrentan (o no se enfrentan, en algunos casos).

Diseño de sujeto único

Amenaza	Historia	Maduración	Selección	Pruebas	Instrumen- tación	Regresión	Mortali- dad	Interferencia de múltiples tratamientos	Disposicio- nes reactivas	Efectos de experi- mentador	Sensibilización por prueba previa
Diseño de grupo control no equivalente	+	+	+	+	+	?	+		?	?	-
Comparación con grupo estático	+	?	-	+	+	+	-			-	
Notas:	+ La amenaza se toma en cuenta. ? Dependiendo del entorno y las circunstancias, las amenazas podrían haberse tomado en cuenta. - Las amenazas no se toman en cuenta. Un espacio en blanco indica que la amenaza no aplica.										

El método experimental, tal como lo describimos en el capítulo anterior; es la forma más común de probar si existen relaciones de causa y efecto, pero no es la única manera.

Existe una estrategia totalmente distinta para entender las diferencias de causa y efecto que no examina grupos, sino individuos. Los **diseños de investigación de sujeto único** son muy comunes en campos como el análisis de la conducta, pero son útiles en casi cualquier entorno en el que un investigador desee conocer los efectos de manipular una variable independiente, sobre la conducta de un individuo. De hecho, podemos decir que mientras que los diseños de grupos (como los que tratan Campbell y Stanley) se concentran en una o más conductas de muchos individuos, los diseños de sujeto único se concentran en muchas conductas (de tipo similar) de un solo individuo. La meta, empero, es la misma: determinar los efectos de una variable independiente sobre la conducta.

No es sólo el método lo que es diferente aquí, sino la forma como se ve la conducta y los aspectos que es importante examinar al realizar la investigación. Los diseños de sujeto único están muy enraizados en la visión conductista del desarrollo, en la que los cambios de comportamiento se ven en función de sus consecuencias. Esta escuela de pensamiento, popularizada por los estudios con animales de B. F. Skinner, ha ayudado a sustentar una visión totalmente distinta de las razones del comportamiento y de la forma como debe estudiarse la conducta. Sin ser mejor o peor que el método de grupos, el método de sujeto único trata de contestar la importante pregunta de causalidad de una forma única y creativa.

El método básico en un experimento de sujeto único es:

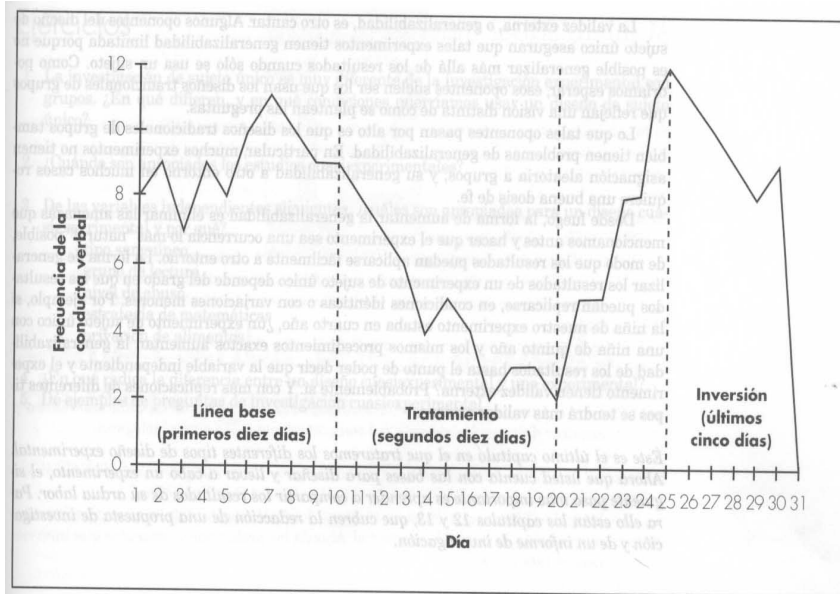
1. Medir una conducta antes del tratamiento.
2. Aplicar un tratamiento.
3. Suspender el tratamiento (inversión).

El supuesto es que si la conducta cambia como resultado del tratamiento, cuando el tratamiento se suspenda la conducta volverá a los niveles que tenía antes del tratamiento. Por tanto, si a un investigador le interesa aumentar el nivel de conducta verbal de un niño especialmente retraído en un grupo escolar, él o ella haría lo siguiente:

1. Medir la tasa de comportamiento verbal cada 5 minutos durante una hora en un periodo de 10 días. El investigador supone que este esquema de medición producirá una muestra representativa de la conducta verbal del niño.

Este primer paso se denomina **línea base**, ya que es contra esta medida que el investigador comparará los resultados del tratamiento para ver si la conducta verbal aumenta.

3. Poner en práctica el tratamiento. Cada vez que el niño exhiba comportamiento verbal, se reforzará con intensos elogios verbales por parte del maestro. Se llevará el mismo tipo de registro durante un periodo de 10 días.
4. Como prueba final de la efectividad del tratamiento, se suspenden los elogios verbales y se mide otra vez la conducta verbal.



En la figura 11.1 se muestra una gráfica del aspecto que podría tener este diseño experimental. Puede verse que la frecuencia del comportamiento disminuyó cuando se aplicó el tratamiento y que aumentó cuando se suspendió el tratamiento.

Figura 11.1 Diseño ABA sencillo, que consiste en establecimiento de una línea Base, tratamiento y medición

En la terminología del diseño de sujeto único, la línea base se rotula A y el tratamiento B. En el ejemplo que acabamos de presentar, el diseño sería del tipo ABA. Como cabría esperar, existen varios diseños que usan las condiciones A y B, siendo el más sencillo el AB, donde se establece una línea base y se pone en práctica un tratamiento. También tenemos una serie de diseños ABAB en los que se alternan las condiciones de línea base y tratamiento

La ventaja primordial del diseño ABAB sobre los diseños AB y ABA es que reduce la preocupación por el problema ético de introducir lo que podría ser un tratamiento correctivo (B) y luego medir sólo los efectos de la suspensión del tratamiento. El diseño ABAB busca reintroducir el tratamiento y hacer que la conducta vuelva al nivel en el que estaba después del primer intento de aplicar el tratamiento. Por ejemplo, si a un niño que tenía una tasa baja de conducta verbal se le deja en esa situación, podrían surgir dudas respecto a si se tomó o no se tomó una decisión ética. Si es fácil usar un diseño ABAB y demostrar una vez más la efectividad de un tratamiento, ¿por qué no hacerlo?

Evaluación de diseños de sujeto único

Podemos aplicar a los diseños de sujeto único los mismos criterios de validez interna y externa como medidas de la confiabilidad que aplicamos a otros diseños.

La mayor parte de los diseños del tipo ABA y ABAB tienen suficiente validez interna. Tales diseños ponen de manifiesto que la manipulación de la variable independiente (su presentación o retiro) hace o no que cambie una conducta. Por tanto, lo que se observa es el resultado de lo que se hizo, y esto es el criterio primordial de la validez interna.

La validez externa, o generalizabilidad, es otro cantar. Algunos oponentes del diseño de sujeto único aseguran que tales experimentos tienen generalizabilidad limitada porque no es posible generalizar más allá de los resultados cuando sólo se usa un sujeto. Como podríamos esperar, esos oponentes suelen ser los que usan los diseños tradicionales de grupos que reflejan una visión distinta de cómo se plantean las preguntas.

Lo que tales oponentes pasan por alto es que los diseños tradicionales de grupos también tienen problemas de generalizabilidad. En particular, muchos experimentos no tienen asignación aleatoria a grupos, y su generalizabilidad a otro entorno en muchos casos requiere una buena dosis de fe.

Desde luego, la forma de aumentar la generalizabilidad es eliminar las amenazas que mencionamos antes y hacer que el experimento sea

una ocurrencia lo más "natural" posible, de modo que los resultados puedan aplicarse fácilmente a otro entorno. La forma de generalizar los resultados de un experimento de sujeto único depende del grado en que los resultados puedan replicarse, en condiciones idénticas o con variaciones menores. Por ejemplo, si la niña de nuestro experimento estaba en cuarto año, ¿un experimento de sujeto único con una niña de quinto año y los mismos procedimientos exactos aumentan la generalizabilidad de los resultados hasta el punto de poder decir que la variable independiente y el experimento tienen validez externa? Probablemente sí. Y con más repeticiones de diferentes tipos se tendrá más validez externa.

Éste es el último capítulo en el que trataremos los diferentes tipos de diseño experimental. Ahora que usted cuenta con las bases para diseñar y llevar a cabo un experimento, el siguiente paso más importante es aprender a compartir los resultados de su ardua labor. Para ello están los capítulos 12 y 13, que cubren la redacción de una propuesta de investigación y de un informe de investigación.

Ejercicios

1. La investigación de sujeto único es muy diferente de la investigación experimental con grupos. ¿En qué difieren, y en qué condiciones querríamos usar un diseño de sujeto único?
2. ¿Cuándo son apropiados los estudios cuasiexperimentales?
3. De las variables independientes siguientes, ¿cuáles son apropiadas para un diseño cuasiexperimental y por qué?
tipo sanguíneo
grupo de lectura
nivel de abuso
estrategia de matemáticas
privación de alimentos
4. ¿En qué radica la diferencia entre un diseño cuasiexperimental y uno experimental?
5. Dé ejemplos de preguntas de investigación cuasiexperimental.

CAPITULO 12 Cómo escribir una propuesta de investigación

En Neil J. Salkind "Métodos de Investigación", Prentice Hall, México, 1998.

Lo que aprenderá en este capítulo

Sólo una cosa: cómo escribir una propuesta

Si uno de los requisitos para este curso es escribir una propuesta de investigación, esta usted en el lugar apropiado. En este capítulo los guiaremos respecto a las consideraciones que debe tener en cuenta al redactar una propuesta para el curso. Incluso si no se le pide redactar una propuesta para el curso, no abandone la lectura. Lo que aprenda aquí le será de utilidad en sus labores de investigación, pues explicaremos que es lo que distingue las propuestas aceptables de las no aceptables. También explicaremos la importancia de plantear una pregunta de manera lógica y clara para que sea fácil contestarla.

Escribir una propuesta no es una tarea fácil para nadie, y podría ser especialmente difícil si no lo ha hecho antes o si no acostumbra escribir mucho. La tarea requiere diligencia, compromiso y mucho trabajo duro. No obstante, todo ese trabajo bien vale la pena, el resultado será un producto del que pueda enorgullecerse y eso es solo el principio. Si realmente persevera y lleva a cabo la investigación propuesta, estará haciendo la contribución importante a su campo. Con estas palabras pasamos a detallar los pasos principales a seguir para redactar una propuesta, comenzando con el aspecto general que tiene.

El formato de una propuesta de investigación

Saber como organizar y presentar una propuesta y es una parte importante del trabajo del trabajo de un investigador. El acto mismo de asentar sus ideas en papel le ayudara a aclarar sus intereses en cuanto a investigación y tener la certeza de estar diciendo lo que quiere decir. Cuanto mas trabajo y reflexión invierta en su propuesta mas fácil le será llevar a cabo la investigación posteriormente.

Con esto en mente he aquí un bosquejo básico de lo que una propuesta de investigación debe contener y unos cuantos comentarios acerca de cada una de estas secciones. Tenga presente que las propuestas se pueden organizar de muchas formas distintas.

I introducción

- A enunciado del problema
 - B justificación de la investigación,
 - 1. planteamiento de los objetivos de la investigación
 - C hipótesis
 - D definiciones de términos
 - E resumen que incluye el replanteamiento del problema
- II (Breve) revisión bibliográfica pertinente
- A importancia de la pregunta que se hace
 - B la situación actual del tema
 - C la relación entre la bibliografía y planteamiento de problemas
- D Resumen que incluye un replanteamiento de las relaciones entre las variables importantes consideradas y la importancia de tales relaciones tiene para la hipótesis propuesta en la introducción
- III Método
- A participantes, incluida una descripción y procedimientos de selección
 - B diseño de la investigación
 - C Planes de recopilación de datos
 - 1 definición operativa de todas las variables
 - 2 confiabilidad y validez de los instrumentos
 - 3 resultados de estudios piloto
 - D análisis propuesto de los datos
 - E resultados de los datos
- IV implicaciones y limitaciones
- V Apéndices
- A copias de los instrumentos que se usaran
 - B resultados de estudios piloto (datos reales)
 - C Aprobación para experimentación con seres humanos
 - D. Forma de permiso de participante
 - E. Línea de tiempo

Si usted ha examinado las tesis o disertaciones de otras personas, tal vez haya notado que este cuadro sinóptico está organizado siguiendo la misma secuencia general de los títulos de los capítulos: introducción, revisión de la bibliografía, metodología, resultados y discusión. Puesto que sólo se trata de una propuesta, las últimas dos secciones no pueden presentar el análisis de los datos reales y comentar los hallazgos. En vez de ello, la propuesta simplemente habla de las implicaciones y limitaciones del estudio y la última parte (V) contiene todos los apéndices importantes.

Una pauta breve acerca de lo que una propuesta debe contener está constituida por las tres primeras secciones de la propuesta terminada: introducción, revisión de la bibliografía y método. El resto del material (implicaciones y demás) se podrá incluir a discreción suya y con base en los deseos de su asesor o profesor. Tenga presente que las primeras tres secciones podrían ser mucho pedir pero de todos modos usted tendrá que recabar esa información, y si lo hace antes de recopilar sus datos adquirirá más confianza en su investigación, además de adelantar el trabajo.

Organícese

Si bien sus palabras son importantes, la apariencia de su propuesta también lo es. Aunque lo que usted dice es más importante que cómo lo dice, mucho hay de verdad en la afirmación de Marshall McLuhan de que el medio es el mensaje. Con eso en mente, he aquí unas cuantas sugerencias sencillas respecto a la preparación de propuestas. Si usted tiene dudas acerca de la presentación (y si su profesor no marcó otras pautas), una buena recomendación es seguir las pautas establecidas en el Manual de Publicaciones de la American Psychological Association (1994). Verá un ejemplo de ese formato en el capítulo 13.

- Todas las páginas deben escribirse a máquina con márgenes de 2.5 cm arriba, abajo, a la izquierda y a la derecha. Esto deja suficiente espacio para comentarios.
- Todas las páginas deberán escribirse a doble espacio.
- Revise todo lo que escriba. Esto no significa usar sólo un revisor ortográfico. Estas maravillas de programas sólo detectan errores de digitalización (escribir una letra por otra), pero dejan pasar algunas faltas de ortografía y no detectan usos incorrectos de las palabras ni faltas de sintaxis. Así, pues, revise su documento dos veces: una vez fijándose en el contenido y otra fijándose en la ortografía y los errores gramaticales.
- El documento final deberá engraparse, sin carpetas ni encuadernaciones elegantes (lo que es demasiado costoso e innecesario).
- Todas las páginas deberán numerarse con una cornisa y un número de página como éstos:

En cuanto al formato del contenido, no hay temor de equivocarse si se sigue el ejemplo que veremos en el capítulo 13, el cual se escribió siguiendo las pautas para la presentación de manuscritos que recomienda la American Psychological Association. Hay algunas diferencias entre lo que describiremos aquí y lo que usted verá en el capítulo 13, pero ninguna es importante. Por ejemplo, las pautas de la APA no piden el nombre del autor en las páginas porque la revisión para las revistas científicas es "ciega". Sería útil para su profesora; en cambio, incluir el nombre de usted en cada página.

Cómo evaluar los estudios que usted lee

Cuando comience a leer artículos de investigación como preparación para leer una propuesta (o simplemente para aprender más acerca del proceso de investigación), será bueno que tenga la seguridad de que puede leer, entender y evaluar el contenido.

Como investigador principiante, es obvio que usted podría no estar capacitado aún para ponerse a evaluar y criticar el trabajo de investigadores reconocidos, ¿verdad? ¡Pues no! Aunque todavía no tenga mucha experiencia en el proceso de investigación, no hay razón para que no pueda leer y evaluar críticamente artículos de investigación. Hasta la investigación más especializada debe escribirse de manera clara y comprensible. Por último, aunque a estas alturas usted no pueda contestar satisfactoriamente todas las preguntas que se listan a continuación, son un buen punto de partida para aprender más. A medida que usted adquiera experiencia, las respuestas irán apareciendo.

Entonces, ¿en qué consiste la buena investigación? Los investigadores Hall, Ward y Comer (1988) se hicieron precisamente esa pregunta acerca de 128 artículos de investigación ya publicados. Después de realizar una encuesta de expertos en investigación, encontraron que las deficiencias siguientes (en orden de aparición) son las críticas más importantes:

- El procedimiento de recopilación de datos no se controló con el debido cuidado.
- Hubo deficiencias en el diseño o plan de la investigación.
- No se indicaron las limitaciones del estudio.
- El diseño de la investigación no concuerda con la pregunta que plantea la investigación.
- El método para seleccionar participantes no fue el apropiado. Los resultados del estudio no se presentaron con claridad.
- Se emplearon métodos equivocados para analizar la información recabada.
- El artículo no está claramente escrito.
- Los supuestos en los que se basa el estudio no están claros.
- Los métodos empleados para realizar el estudio no se describieron claramente o ni siquiera se describieron.

¡Vaya serie de peligros que acechan! Para ayudarle a evitar los peores, he aquí un conjunto de preguntas que conviene hacer acerca de cualquier artículo de investigación.

Criterios para Juzgar un estudio de investigación

La revisión de investigaciones previas

1. La bibliografía revisada para el estudio, ¿qué tan íntimamente está relacionada con los trabajos realizados anteriormente en el campo?
2. ¿Es reciente la revisión? ¿Hay referencias sobresalientes que usted conozca y que se omitieron?

El problema y el propósito

3. ¿Puede usted entender el planteamiento del problema?
4. ¿Se expresó claramente el propósito del estudio?
5. ¿Ese propósito está evidentemente vinculado con la bibliografía que se revisó?
6. ¿Se expresó claramente el objetivo del estudio?
7. ¿Existe una justificación conceptual para las hipótesis?
8. ¿Se presenta una justificación de la importancia del estudio.

La hipótesis

9. ¿Se expresaron con claridad las hipótesis de investigación?
10. ¿Se plantearon explícitamente las hipótesis de investigación?
11. ¿Las hipótesis expresan una asociación clara entre variables?
12. ¿Las hipótesis están fundamentadas en una teoría o en una revisión y presentación de bibliografía pertinente?
13. ¿Las hipótesis son susceptibles de probarse?

El método

14. Se definieron claramente las variables tanto independientes como dependientes?

15. ¿Es completa la definición y la descripción de las variables?
16. ¿Queda clara la forma en que se realizó el estudio?

La muestra

17. ¿La muestra se seleccionó de tal manera que parece ser representativa de la población?
18. ¿Es evidente de dónde provino la muestra y cómo se seleccionó?
19. ¿Qué tan parecidos son los sujetos del estudio a los que se han usado en otros estudios similares?

Resultados y discusión

20. ¿El autor relaciona los resultados con la revisión de la bibliografía?
21. ¿Los resultados están relacionados con la hipótesis?
22. ¿El análisis de los resultados es congruente con los resultados?
23. ¿El análisis permite tomar una decisión definitiva en un sentido o en otro respecto a la hipótesis que presentó el autor?

Referencias

24. ¿Está actualizada la lista de referencias?
25. ¿Las referencias son consistentes en cuanto a su formato?
26. ¿Están completas las referencias?
27. ¿La lista de referencias refleja algunas de las fuentes de referencias más importantes en el campo?
28. ¿Todas y cada una de las referencias citadas en el cuerpo del artículo aparece en la lista de referencias?

Comentarios generales acerca del informe

29. ¿Está escrito con claridad y es comprensible?
30. ¿El lenguaje empleado carece de predisposiciones (no es sexista y relativamente independiente de la cultura)?
31. ¿Cuáles son los puntos fuertes y débiles de la investigación?
32. ¿Cuáles son las implicaciones primarias de la investigación?
33. ¿Qué haría usted para mejorar la investigación?

Planificación de la investigación real

Ya está usted en buen camino para formular hipótesis aceptables con las que se puede trabajar, y ahora sabe al menos cómo comenzar a revisar la bibliografía y sacar algo en claro de los cientos de recursos disponibles. Sin embargo, lo que tal vez no sepa, sobre todo si nunca ha participado en algún tipo de trabajo de investigación, es cuánto tiempo va a necesitar para sus actividades, desde la primera visita a la biblioteca hasta su examen final o la presentación del informe de investigación terminado. Eso es lo que aprenderá aquí.

Aunque todavía le falta mucho por aprender acerca del proceso de investigación, ahora es un buen momento para darse una idea de qué otras actividades tendrá que realizar para poder llevar a cabo su proyecto de investigación. También es recomendable hacer una estimación de cuánto tiempo podrían tomar dichas actividades.

Primero las actividades. En la tabla 12.1 se muestra un ejemplo de lista de verificación de actividades que es preciso realizar para preparar la propuesta (o efectuar la investigación). Las actividades se han agrupado bajo los títulos generales que presentamos antes.

Muy bien, ahora pasemos al cálculo de cuánto tiempo vamos a necesitar. Una forma efectiva de hacerlo es estimar cuánto tiempo le tomará cada actividad individual (escribir la revisión de la bibliografía, recopilar los datos, etc.) utilizando alguna medida estándar como días y considerando que a veces las cosas

- resultan tal como se planearon,
- no resultan tan bien como se planearon, o
- resultan mal (lo cual es la regla más que la excepción).

Calcule ahora el promedio de estos valores. En términos más precisos, divida los días hábiles en secciones de 4 horas (matutina y vespertina); cada sección será una unidad de tiempo. Entonces, una semana consiste en 10 unidades de tiempo.

Por ejemplo, consideremos la actividad de *Búsqueda en fuentes primarias* (como parte de la revisión de la bibliografía) y estimemos que nos tomará

- cuatro días, u ocho unidades de tiempo, si todo sale perfecto,
- seis días, o 12 unidades, si las cosas no resultan exactamente como las planeamos, y
- ocho largos días, o 16 unidades, si las cosas resultan mal.

Una vez que tenga las estimaciones anteriores, simplemente calcule el promedio para la actividad, y tendrá una sola estimación de cuánto deberá tardar la actividad en cuestión, digamos

$$(8 + 12 + 16) / 3 = 12 \text{ unidades}$$

o seis días, que podría ser una semana saturada de trabajo (si trabaja el sábado o el domingo).

Si quiere ser aún más preciso, pondere las estimaciones. Por ejemplo, digamos que espera tener muchos problemas para encontrar una muestra, y en el mejor de los casos podría esperar que las cosas no salgan muy mal. En cambio, redactar la sección descriptiva deberá ser fácil. Entonces, pondere la estimación "cuando hay un poco de atraso" multiplicándola por un factor de 2 o 3.

Lista de verificación para planear la redacción de la propuesta y las actividades de investigación				
Dentro de cada celda anote las estimaciones de cuántas unidades (o medios días) tomará cada actividad. Si va a escribir una propuesta considere sólo las actividades relacionadas con la propuesta. Si va a escribir la propuesta y va a realizar la investigación, considere todas las actividades. Sea justo consigo mismo y también razonable. No eleve tanto las expectativas que no pueda cumplir; ni se dé tanto tiempo que nunca termine				
	Actividad	Estimación cuando todo va según el plan	Estimación cuando hay un poco de atraso	Estimación cuando hay un atraso considerable
Introducción	<ul style="list-style-type: none"> • Examinar fuentes generales hasta que se le ocurra una idea • Formular una pregunta de investigación • Presentar una hipótesis preliminar 			
Revisión de la bibliografía	<ul style="list-style-type: none"> • Buscar fuentes secundarias • Buscar fuentes primarias • Reconsiderar la bibliografía y plantear la hipótesis de investigación 			
Metodología	<ul style="list-style-type: none"> • Identificar y describir las variables independientes • Identificar y describir las variables dependientes • Someter a pruebas de campo las variables dependientes • Crear formatos de captura de datos • Localizar una muestra adecuada • Someter la hipótesis de investigación a una prueba piloto • Distribuir los formatos de permiso 			

	<ul style="list-style-type: none"> • Recopilar datos 			
Resultados	<ul style="list-style-type: none"> • Analizar los datos • Informar los resultados usando tablas y gráficas si es necesario 			
Discusión	<ul style="list-style-type: none"> • Revisar la naturaleza y el propósito de la investigación • Referirse a los resultados a la luz de la investigación • Sacar la conclusión apropiada acerca de la confirmación o refutación de la hipótesis de investigación • Analizar las limitaciones del estudio • Analizar las implicaciones del estudio • Comentar las direcciones futuras 			

Puede calcular este tipo de estimaciones para todas las actividades que se indican en la tabla 12.1 y luego sumarías para obtener una estimación para la actividad global. tenga presente que todo tarda más de lo que usted cree, así que sea generoso hasta en sus estimaciones más optimistas.

Selección de la variable dependiente

En varios puntos de este libro hemos mencionado la importancia de seleccionar una variable dependiente o una medida del resultado con extremo cuidado. Esta variable es el vínculo entre toda la laboriosa preparación y reflexión que ha realizado y la conducta real que usted desea medir. Incluso si tiene una idea magnífica para un proyecto de investigación y su hipótesis da justo en el blanco, una mala selección de la variable dependiente echará todo por tierra.

He aquí nueve cosas que debemos recordar al seleccionar dicha variable. Utilícelas como lista de verificación al examinar estudios anteriores en busca de lo que necesita.

- Trate de encontrar medidas que se hayan usado antes. Esto les confiere credibilidad y le permite a usted justificar su decisión citando usos previos en otros estudios de investigación.
- Asegúrese de que ya se haya establecido la validez de la medida. No es recomendable seleccionar variables dependientes cuya validez todavía no se ha establecido o es baja. Hacerlo haría surgir demasiadas dudas acerca de la integridad de todo el estudio. Recuerde que puede averiguar si se ha demostrado la validez de una prueba haciendo una revisión de otros estudios en los que se ha usado esa prueba, o examinando los manuales que acompañan a la prueba o herramienta de evaluación.
- Asegúrese de que ya se haya establecido la confiabilidad de la medida. Al igual que la validez, la confiabilidad es una característica crucial de las variables dependientes útiles.
- Si la prueba requiere capacitación especial, considere el tiempo y la dedicación que implica aprender a usarla. No estamos hablando de limitarse a leer las instrucciones y practicar la administración de una prueba. Estamos hablando de una capacitación intensiva como la que se requiere para administrar pruebas de inteligencia y diversas escalas de personalidad.
- Asegúrese de poder obtener un ejemplar de la prueba antes de tomar una decisión respecto a si usarla o no. Es posible que usted haya leído acerca de ella en un estudio previo, pero en tanto no examine sus pautas en la población con la que se va a usar, los requisitos para la administración, los costos, etc., no es conveniente que tome una decisión definitiva. Por lo regular es posible obtener un paquete de muestra sin costo o por una cuota mínima del creador o editor de la prueba (aunque es posible que para ello necesite una carta de su asesor, ya que varias compañías que publican pruebas no envían materiales a cualquiera que los solicita.)

- Si va a necesitar normas, asegúrese de que estén disponibles. Algunas pruebas no requieren normas, pero si su intención es comparar el desempeño de diferentes muestras con los puntajes de una población más general, hay que tener algo con qué hacer la comparación. Como veremos más adelante, las normas son especialmente importantes en el caso de pruebas referidas a normas.
- Asegúrese de obtener la versión más moderna de la prueba. Los editores constantemente modifican los materiales de prueba, sea cambiando su presentación o modificando los datos normativos o de confiabilidad y validez mismos. Simplemente pregunte: "¿Es ésta la versión más reciente que hay?"
- La prueba tiene que ser apropiada para el grupo de edades con que va a trabajar. Si una prueba mide algo en sujetos de 10 años de edad, esto no implica que es igualmente confiable y válida con sujetos de 20 años, o siquiera que medirá el mismo constructo o conducta subyacente a esa edad. Busque otras formas de la misma prueba u otra prueba que mida el mismo constructo en el grupo de edades de interés.
- Por último, busque reseñas de la prueba en las distintas revistas científicas y fuentes de referencias como *Tests in Print* (Mitchell, 1983), que lista cientos de pruebas de todo tipo, y *Mental Measurement Yearbook* (Conoley y Kramer, 1989). Ambas publicaciones contienen amplia información acerca de diferentes tipos de prueba que incluyen todo desde los procedimientos de administración hasta reseñas críticas de las pruebas escritas por expertos externos. Estudie estas reseñas críticas antes de decidir adoptar un instrumento.

Revisión de pruebas

Todavía hay más acerca de la selección de variables dependientes (o de medidas de preselección para realizar la asignación a grupos que fungirá como variable independiente). En el mejor de los casos, cuando todo parece marchar viento en popa, es difícil encontrar la prueba exacta que uno quiere usar para diagnosticar, evaluar, determinar efectos, usar como herramienta de colocación, etc. La variable dependiente que usted seleccione podría incluso no ser una *prueba* en el sentido formal de la palabra. Sin embargo, si se trata realmente de una prueba, es preciso considerar muchas características y cualidades del instrumento.

Con eso en mente, presentaremos una relación de los criterios que debemos considerar y que permiten comparar y contrastar diferentes pruebas. Para cada prueba que desee considerar, obtenga la información indicada hasta donde sea posible y luego utilícela para tomar una decisión. No olvide ponderar cada uno de los criterios de acuerdo con las circunstancias específicas. Por ejemplo, aunque una prueba sea exactamente lo que se necesita en cuanto a su diseño y propósito, si su costo es prohibitivo o si se necesita capacitación especial (que usted no posee) para administrarla, es poco probable que pueda usarla.

Información básica

1. Nombre de la prueba:
2. Fecha de publicación:
3. Autor(es) de la prueba:
4. Editorial:
5. Costo de todos los materiales de prueba necesarios:
6. Costo del paquete de muestra:

Información general sobre la prueba

7. Propósito de la prueba según su(s) autor(es):
8. Propósito de la prueba según su uso en otros estudios:
9. Niveles de edad incluidos:
10. Años escolares incluidos:
11. Poblaciones especiales incluidas:
12. Método de administración (individual o por grupos):
13. Método de calificación (manual o por computadora):
14. Tiempo de administración:
15. Facilidad de administración:
16. Facilidad de calificación:
17. Capacitación requerida para administrarla:
18. Calidad del manual de la prueba y otros materiales:

Diseño y apariencia

19. Instrucciones claras y directas:
20. Diseño y producción satisfactorios:
21. Organización de los reactivos en la página:
22. Facilidad de lectura:

Confiabilidad

23. Datos de confiabilidad que se proporcionan:

24. Tipo de confiabilidad establecida (prueba original y repetida, formas paralelas, etc.):
25. Estudios independientes realizados para establecer la confiabilidad:

Validez

26. Datos de validez que se proporcionan
27. Tipo de validez establecida
28. Estudios independientes realizados para establecer la validez

Normas

29. Normas disponibles:
30. Descripción de los grupos de norma:
31. Cómo se seleccionaron los grupos de norma:
32. Idoneidad de los grupos de norma para sus propósitos:

Evaluación

33. Cómo se usó en el pasado:
34. Resumen de reseñas externas:
35. Otra información de evaluación:

Selección de una muestra

Muchos investigadores piensan que no hay nada más importante que seleccionar una muestra que refleje con exactitud las características de la población que les interesa estudiar. Sin embargo, la selección de muestras a veces puede ser un asunto riesgoso, y es preciso contestar todo tipo de preguntas antes de poder iniciar el proceso de selección de muestras. He aquí una lista de factores que hay que tener presentes.

- Piense que está tratando de encontrar una reserva adecuada de candidatos de entre los cuales seleccionar una muestra, y multiplique por 100 el número de otras personas que están tratando de hacer lo mismo en su comunidad. Ésa es una estimación conservadora de cuántas personas en una comunidad universitaria están buscando

una muestra que incluir en su estudio. ¿Dónde puede usted buscar? Puede intentarlo en los siguientes lugares:

Grupos de iglesias y Sinagogas
Niños y niñas exploradores
Asilos y comunidades de ancianos
Centros preescolares
Clubes de solteros
Grupos de interés especial y de aficionados
Organizaciones fraternales

Recuerde que no conviene seleccionar un grupo que está organizado por alguna razón específica si dicha razón está relacionada, aunque sea remotamente, a lo que usted está estudiando. No seleccionaría, por ejemplo, miembros de una logia masónica si lo que desea estudiar es la lealtad y la amistad, así como no seleccionaría padres que envían a sus hijos a escuelas privadas si desea realizar encuestas acerca del apoyo a la educación pública, a menos que la Selección de tales muestras sea parte importante de su plan de muestreo.

· Acérquese a los candidatos con una idea absolutamente clara de qué es lo que quiere hacer; cómo lo quiere hacer y qué obtendrán ellos a cambio (un curso gratuito, los resultados del estudio o cualquier otra cosa que usted crea podría serles de beneficio).

· la población debe tener las características de los grupos que usted desea estudiar. Sobra decir (aunque lo diremos de todos modos) que seleccionar una muestra de una población que no se ha identificado claramente es el primer error grave en la selección de muestras. Si quiere estudiar niños en edad preescolar, no puede estudiar niños de primero de primaria simplemente porque los dos grupos están cercanos en cuanto a sus edades. La experiencia preescolar y la de primer año son muy diferentes.

· El tipo de investigación que realice dependerá del tipo y tamaño de la muestra que necesite. Por ejemplo, si va a realizar investigación descriptiva mediante estudios de casos, lo que implica largas e intensas entrevistas y tiene generalizabilidad limitada (lo cual no es uno de los fines del método), necesitará muy pocos sujetos en su muestra. Si va a realizar un estudio de diferencias entre grupos, necesitará al menos 30 para cada grupo.

· Una prueba muy confiable producirá resultados más exactos que un examen de ensayo hecho en casa. Cuanto menos confiables y válidos sean sus instrumentos, más grande será la muestra que necesitará para obtener una imagen exacta de lo que desea estudiar.

· Considere los recursos financieros de que dispone. Cuanto más dinero (y recursos en general) tenga, podrá aplicar la prueba a más sujetos. Recuerde que cuanto mayor sea la muestra (hasta cierto punto), mejor; porque las muestras grandes aproximan mejor la población de la que forman parte.

El número de variables que esté estudiando y el número de grupos que utilice afectarán el proceso de Selección de la muestra. Si simplemente está examinando la diferencia en habilidades verbales entre hombres y mujeres, puede bastar con 25 030 sujetos en cada grupo. Si añade la edad (5 y 10 años) y la clase social (alta y baja), tendrá hasta ocho posibles combinaciones distintas (como niñas de 5 años de clase social alta) y hasta $8 \times 30 = 240$ sujetos en una muestra de tamaño adecuado.

Recopilación y análisis de los datos

Si usted ha Seguido los distintos pasos de este capítulo, es probable que ya haya llevado a cabo, esté planeando efectuar o pueda hacer lo siguiente:

- Ha entendido el formato de las propuestas de investigación.
- Ha escogido un problema de cierta importancia en su campo y ha especificado las variables de interés (tanto dependientes como independientes).
- Ha encontrado medidas confiables y válidas de la variable dependiente.

Ahora está usted listo para iniciar la etapa de análisis de los datos.

En los capítulos 7 y 8 usted aprendió a usar algunas herramientas estadísticas válidas para describir las características de los datos que recopilará durante las primeras etapas de su investigación.

En este punto de su propuesta es preciso realizar las siguientes tareas. No continúe antes de haberlas llevado a cabo.

1. Creación de un formato de recopilación de datos que le ayude a organizarse y obtener resultados exactos.
2. Especificación del tipo de estadística descriptiva que usará para reseñar las variables que está examinando. ¿En qué nivel de medición están, y cuál nivel nominal, ordinal, de intervalo, de razón- es el que mejor refleja lo que usted está tratando de decir?
3. Identificación de los otros tipos de información que usted necesita presentar en este análisis inicial de la naturaleza de sus datos. Tal vez necesita información demográfica, como el género de los participantes, su edad, clase social o afiliación política. Incluso si esta información no está relacionada directamente con la pregunta que usted está haciendo, no está de más recopilarla en este punto. Quizá después se dé cuenta de que le conviene regresar y examinar esa otra información, y se alegrará de haberla tenido. Esto no implica que debe usar un cuestionario demográfico de 10 páginas con más de 1000 preguntas. Lo que implica es que, dentro de lo razonable, hay que recabar información relacionada aunque sea indirectamente con la pregunta principal.
4. Recopilación piloto de datos, para poder practicar las técnicas de estadística descriptiva e inferencial que vimos en los capítulos 7 y 8. Trate el análisis como si filera el definitivo y siga todos y cada uno de los pasos que piense seguir en el análisis final de los datos. De este modo usted sabrá exactamente qué entiende y qué no comprende bien, y podrá obtener ayuda si es necesario. Realice el análisis de los datos tanto a mano, utilizando las fórmulas de este libro, como con SPSS, que veremos en el apéndice A.

Selección de una estadística inferencia

Por más experiencia que usted tenga, la selección de una prueba inferencial siempre es una tarea que requiere cuidado. Si usted apenas se inicia en investigación, como es el caso, la decisión puede ser intimidante.

El diagrama de flujo de la figura 12.2 indica qué hacer en algunas de las situaciones más comunes, como probar la diferencia entre las medias de dos o más grupos y examinar las relaciones entre grupos. En ambos casos se aplican los mismos principios en cuanto a determinar la significatividad de un resultado.

No piense ni por un momento que (a) es posible sustituir un diagrama como el de la figura por un curso de estadística básica, o que (b) éste es un curso de estadística básica. El diagrama sólo pretende ser una ayuda sencilla para guiarlo hacia una selección correcta. Explicamos un poco el *por qué* de la inferencia en el capítulo 8, pero para entender a fondo las razones inscribese en un curso de estadística básica y haga feliz a su asesor.

Protección de sujetos humanos

Como explicamos en el capítulo 2, la mayor parte de las organizaciones que patrocinan investigaciones (como las universidades) tiene algún tipo de comité que con regularidad revisa las propuestas de investigación para garantizar que las personas (y animales) no corran peligro alguno si llegan a participar.

Antes de que los investigadores inicien su trabajo, y como parte del proceso de preparar la propuesta, habrán llenado y anexado a la propuesta un formato como el que presentamos en la figura 2.4. El comité revisará la información y aprobará el proyecto (indicando que los sujetos humanos no están en peligro) o bien tratará de colaborar con el investigador en la modificación de los métodos propuestos con el fin de que las cosas se lleven a cabo según lo planeado.

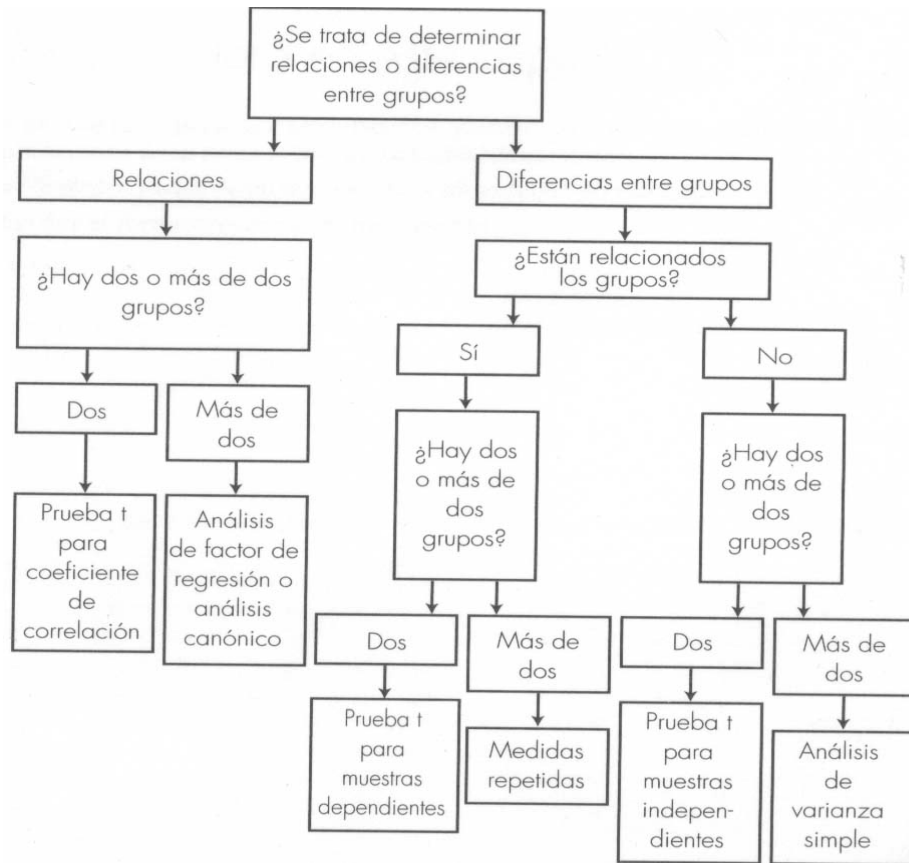


Figura 12.1 Selección de un prueba estadística: ejemplo para comparar grupos.

Los científicos con éxito le dirán que si parte de una pregunta clara y bien meditada, el resto de su propuesta, así como la ejecución de su investigación, se facilitará muchísimo.

Por otro lado, si su pregunta inicial es poco clara, pronto se verá en problemas y será incapaz de concentrarse en el verdadero objetivo del proyecto. trabaje todos los días en la redacción de su propuesta, léala una y otra vez, déjela reposar un tiempo, haga que un amigo o colega le de un vistazo y ofrezca sugerencias, escriba otro poco más, dejela reposar otro rato... ¿Capta el mensaje? Practique y trabaje duro, y será bien recompensado.

Ejercicios

1. Vaya a la biblioteca y seleccione un artículo de una *revista* científica que represente un trabajo en su campo de interés. Aplique cada uno de los criterios que especificamos en este capítulo (vea la sección "Criterios para juzgar un estudio de investigación"). Para que este ejercicio sea todavía más interesante, realícelo junto con un colega, o escoja el mismo artículo que un colega y compare sus resultados con los de él.
2. Llene el formato para experimentación con seres humanos de la figura 2.4. Si está trabajando en un proyecto de investigación, incluya la información real. Si no, incluya información hipotética.