relaciones y la prueba de hipótesis. En otras palabras, la entrevista es un instrumento de medición psicológica y sociológica. Quizás más preciso, los productos de entrevistas —las respuestas de los entrevistados a preguntas cuidadosamente elaboradas— pueden traducirse en medidas de variables. Por lo tanto, las entrevistas y los inventarios de entrevista están sujetos a los mismos criterios de confiabilidad, validez y objetividad que otros instrumentos de medición.

Una entrevista sirve para tres propósitos principales:

1. Como un dispositivo exploratorio para ayudar a identificar variables y relaciones, para sugerir hipótesis y para guiar otras fases de la investigación.
2. Ser el principal instrumento de la investigación. En dicho caso, en el inventario de entrevista se incluyen preguntas diseñadas para medir las variables de la investigación. Estas preguntas se consideran después como reactivos en un instrumento de medición, más que como meros dispositivos para reunir información.
3. Puede complementar otros métodos: hacer un seguimiento de resultados inesperados, validar otros métodos y profundizar en las motivaciones de entrevistados y en las razones por las que responden como lo hacen.

Al utilizar entrevistas como herramientas de investigación científica, debe plantearse la pregunta: ¿los datos del problema de investigación pueden obtenerse de una manera mejor y más fácil? Lograr confiabilidad, por ejemplo, no constituye un problema pequeño. Los entrevistadores deben tener un entrenamiento; las preguntas deben probarse con anterioridad y revisarse para eliminar ambigüedades y redacción inadecuada. ¿Vale la pena el esfuerzo? Tampoco la validez es un problema pequeño. Deben realizarse esfuerzos especiales para eliminar los prejuicios del entrevistador; las preguntas deben probarse para encontrar sesgos desconocidos. El problema de investigación particular y la naturaleza de la información buscada debe, en el análisis final, determinar si se utilizará o no la entrevista. Cannell y Kahn (1968, capítulo 15) y Warwick y Lininger (1975, capítulo 7) proporcionan una guía detallada respecto a si una entrevista debe o no ser empleada.

La entrevista

La *entrevista* es una situación interpersonal cara a cara donde una persona (el entrevistador) le plantea a otra persona (el entrevistado) preguntas diseñadas para obtener respuestas pertinentes al problema de investigación. Existen dos tipos generales de entrevista: la *estructurada* y la *no estructurada*, o *estandarizada* y *no estandarizada* (véase Cannell y Kahn, 1968). En la entrevista estandarizada las preguntas, su secuencia y su redacción son fijas. Se permite cierta libertad al entrevistador al plantear las preguntas, pero ésta es relativamente poca. El *Manual del Entrevistador* (1976) producido por el Instituto de Investigación Social (Institute for Social Research) de la Universidad de Michigan establece que la perspectiva de la entrevista está evolucionando de la perspectiva tradicional. La entrevista se considera como una interacción, una relación de roles activos entre entrevistador y entrevistado, donde el entrevistador es incluso un maestro. Cannell y Kahn (1968) y Dohrenwend y Richardson (1963) ofrecen información adicional sobre este tema. El nivel de libertad se especifica de antemano. Las entrevistas estandarizadas utilizan inventarios de entrevista que se han preparado cuidadosamente para obtener información concerniente al problema de investigación.

Las entrevistas no estandarizadas son más flexibles y abiertas. A pesar de que los propósitos de investigación determinan las preguntas planteadas, su contenido, secuencia y

redacción están en manos del entrevistador. Por lo común no utilizan ningún inventario. En otras palabras, la entrevista no estandarizada y no estructurada constituye una situación abierta, en contraste con la entrevista estandarizada y estructurada, que es una situación cerrada. Ello no significa que una entrevista no estandarizada sea casual; debe ser planeada tan cuidadosamente como la estandarizada. Green y Tull (1988) afirman que las entrevistas no estructuradas obtienen información que las entrevistas estructuradas no ofrecen. Con el modelo informal de la entrevista no estructurada el investigador obtiene ideas respecto a las motivaciones del entrevistado. Las entrevistas no estructuradas algunas veces se denominan *entrevistas profundas*. Son especialmente útiles para realizar estudios exploratorios. El interés central aquí lo representa la entrevista estandarizada. Sin embargo, se reconoce que muchos problemas de investigación pueden requerir, y muchas veces es así, un tipo de entrevista en que el entrevistador tenga permitido utilizar preguntas alternativas que se ajusten a entrevistados particulares y a cuestiones particulares. El procedimiento real de la conducción de una entrevista no se analiza en este libro. El lector encontrará una guía en la sección de sugerencias de estudio del presente capítulo.

El inventario de entrevista

La conducta de entrevistar es, en sí misma, un arte, pero la planeación y la redacción de un inventario de entrevista lo es todavía más. Es poco común que un novato produzca un buen inventario, al menos sin considerable estudio y práctica previos. Existen varias razones para ello; las principales probablemente sean el significado múltiple y la ambigüedad de las palabras, la ausencia de un enfoque directo y constante en los problemas e hipótesis de estudio, la falta de apreciación del inventario como un instrumento de medición y la ausencia de antecedentes y experiencia necesarios.

Tipos de información y reactivos de los inventarios

La mayor parte de los inventarios incluyen tres tipos de información: información de la carátula (identificación), información de tipo censal (o sociológica) e información del problema. Con excepción de la identificación, estos tipos de información se analizaron en un capítulo anterior. No obstante, debe mencionarse la importancia de identificar cada inventario de manera precisa y completa. El investigador cuidadoso debe aprender a identificar con letras, números u otros símbolos cada inventario y cada escala. Además, debe registrarse de forma sistemática la identificación de la información de cada individuo. Existen dos tipos de reactivos de inventario que se usan comúnmente: *de alternativa fija* (o cerrados) y *de preguntas abiertas*. Se utiliza también un tercer tipo de reactivo, de alternativas fijas: los reactivos *de escala*.

Reactivos de alternativa fija

Como su nombre lo indica, los reactivos de alternativa fija ofrecen al entrevistado una opción entre dos o más alternativas. A estos reactivos también se les llama preguntas *cerradas* o *de encuesta*. El tipo más común de reactivo de alternativa fija es dicotómico: plantea preguntas que pueden responderse como *sí* o *no*, *de acuerdo* o *en desacuerdo* y otro tipo de respuestas de dos opciones. Con frecuencia se añade una tercera posibilidad: *no sé* o *indeciso*.

Un ejemplo de un reactivo de alternativa fija sería:

¿Considera usted que el gobierno de Estados Unidos ya encontró una cura para el SIDA, pero que la está ocultando?

Sí ..... [ ]
No ..... [ ]
No sé ..... [ ]

A pesar de que los reactivos de alternativa fija poseen las claras ventajas de lograr una mayor uniformidad de la medición y, por lo tanto, mayor confiabilidad, de forzar al entrevistado a responder en una forma que se ajuste a las categorías previamente establecidas y de ser fáciles de codificar, también tienen ciertas desventajas. La mayor de ellas es su superficialidad: sin sondeo, generalmente no van más allá de la superficie de la respuesta. También pueden irritar al entrevistado que no encuentre ninguna alternativa adecuada para él. Peor aún, es posible que fuercen respuestas. Un entrevistado puede elegir una alternativa para ocultar ignorancia o elegir alternativas que no representan con precisión los hechos u opiniones. Tales dificultades no implican que los reactivos de alternativa fija sean malos o inútiles. Por el contrario, se utilizan con buenos resultados si se redactan de manera juiciosa, si se utilizan con un sondeo y si se mezclan con reactivos abiertos. Un *sondeo* es un dispositivo utilizado para encontrar información de los entrevistados sobre un tema, sus marcos de referencia o, más común, para aclarar y establecer las razones de las respuestas dadas. El sondeo incrementa el poder de "obtención de respuesta" de las preguntas, sin cambiar su contenido. Ejemplos de sondeo son: "Dígame más sobre esto." "¿Cómo es eso?" "¿Puede explicarlo?" (véase Warwick y Lininger, 1975, pp. 210-215).

Reactivos abiertos

Los reactivos abiertos o de final abierto representan un desarrollo extremadamente importante en la técnica de la entrevista. Las preguntas *abiertas* son aquellas que brindan un marco de referencia para las respuestas de los entrevistados, pero poniendo un mínimo de restricción a las respuestas y a su expresión. Aunque su contenido está determinado por el problema de investigación, no imponen ninguna otra restricción sobre el contenido ni forma de las respuestas del entrevistado. Más tarde se darán ejemplos.

Las preguntas abiertas tienen importantes ventajas, aunque también tienen desventajas. Sin embargo, si se redactan y utilizan apropiadamente se minimizan las desventajas. Las preguntas abiertas son flexibles; tienen la posibilidad de profundizar; le permiten al entrevistador aclarar malos entendidos (a través del sondeo), establecer la falta de conocimiento de un entrevistado, detectar ambigüedades, promover la cooperación y lograr *rapport* y mejores estimados de las verdaderas intenciones, creencias y actitudes de los entrevistados. Su empleo tiene también otra ventaja: en ocasiones las respuestas a preguntas abiertas *sugieren* posibilidades de relaciones e hipótesis. Los entrevistados algunas veces darán respuestas inesperadas que tal vez indiquen la existencia de relaciones no anticipadas originalmente.

Un tipo especial de pregunta abierta es la pregunta de *embudo*. En realidad se trata de un conjunto de preguntas dirigidas a obtener información sobre un solo tema importante o sobre un solo conjunto de temas relacionados. El embudo inicia con una pregunta general y se va reduciendo progresivamente al punto o puntos específicos importantes. Warwick y Lininger (1975) señalan que el mérito del embudo es que permite la libre respuesta en las primeras preguntas y después se reduce a preguntas y respuestas específicas; y que también facilita el descubrimiento de los marcos de referencia del entrevistado. Otra forma de pregunta de embudo inicia con una pregunta abierta general y continúa con reactivos específicos cerrados. La mejor forma de reconocer las buenas preguntas abiertas y de embudo es por medio del estudio de ejemplos.

Para obtener información sobre prácticas de crianza de niños, Sears, Maccoby y Levin (1957) utilizaron varias buenas preguntas abiertas y de embudo. Una de ellas, con comentarios del autor entre corchetes, es:

Ejemplo

> Por supuesto que todos los bebés lloran. [Note que el entrevistador tranquiliza al padre respecto al llanto de su hijo.] Algunas madres consideran que si se levanta a un bebé cada vez que llora, se le malcría. Otros piensan que no se debe dejar llorar demasiado tiempo a un bebé. [El marco de referencia se ha expresado claramente. También se ha tranquilizado a la madre pues no importa cómo maneje el llanto de su bebé.] ¿Qué piensa usted respecto a esto?
>
> (a) ¿Qué hizo con X respecto a esto?
>
> (b) ¿Qué hizo a la medianoche?

Este conjunto de preguntas de embudo no sólo evalúa actitudes, sino que también sondea prácticas específicas.

Reactivos de escala

Un tercer tipo de reactivo de inventario es el reactivo de escala. Una *escala* es un conjunto de reactivos verbales, a cada uno de los cuales un individuo responde expresando grados de acuerdo o desacuerdo, o algún otro modo de respuesta. Los reactivos de escala tienen alternativas fijas y colocan al individuo entrevistado en algún punto de la escala. (Serán analizados con mayor profundidad en el capítulo 30.) El uso de reactivos de escala en inventarios de entrevista es un desarrollo que promete mucho, debido a que se combinan los beneficios de las escalas con los de las entrevistas. Por ejemplo, se puede incluir una escala para medir actitudes hacia la educación en un inventario de entrevista sobre el mismo tema. Las puntuaciones de la escala se obtienen de esta forma para cada entrevistado y se verifican contra datos de preguntas abiertas. Es posible medir la *tolerancia hacia la inconformidad*, como lo hizo Stouffer (1955), al tener una escala para medir esta variable incluida en el inventario de entrevista.

Criterios para la redacción de preguntas

Los criterios o preceptos para la redacción de preguntas se han desarrollado a través de la experiencia y la investigación. Algunos de los más importantes se presentan más adelante en forma de preguntas. Se anexaron breves comentarios a las preguntas. Al enfrentarse con la necesidad real de redactar un inventario, el estudiante deberá consultar tratados más extensos, ya que el siguiente análisis, en congruencia con la postura del resto del capítulo, pretende ser sólo una introducción al tema. Si se desea una guía práctica véase Emory (1976, capítulo 8), quien proporciona un buen resumen y puntos clave para elaborar el inventario; así como Noelle-Neuman (1970) y Warwick y Lininger (1975). Emory enfatiza la forma de probar el instrumento antes de usarlo realmente, cómo secuenciar los reactivos o preguntas y qué hacer bajo ciertas situaciones. Otras recomendaciones son Atkinson (1971), Beed y Stimson (1985) y Mishler (1986).

1. *¿Está relacionada la pregunta con el problema y los objetivos de investigación?* Con excepción de preguntas de información factual y sociológica, todos los reactivos del inventario deben estar en función del problema de investigación. Esto significa que el

propósito de cada pregunta es generar información que sirva para probar las hipótesis de la investigación.

2. *¿Es apropiado el tipo de pregunta?* Alguna información puede obtenerse mejor con preguntas abiertas —razones para comportamientos, intenciones y actitudes—. Por otro lado, otro tipo de información puede obtenerse de forma más expedita por medio de preguntas cerradas. Si todo lo que se requiere de un entrevistado es la opción preferida de dos o más alternativas, y estas alternativas pueden especificarse con claridad, sería inútil utilizar una pregunta abierta (véase Dohrenwend y Richardson, 1963; Schuman y Presser, 1979; Warwick y Lininger, 1975).
3. *¿El reactivo es claro y sin ambigüedades?* Un reactivo o afirmación ambigua es aquella que permite o invita a interpretaciones alternativas, de las cuales resultan respuestas diferentes. Las preguntas denominadas de doble sentido, por ejemplo, son ambiguas debido a que proporcionan dos o más marcos de referencia en lugar de uno solo. Los entrevistados, aun cuando no se confundan con la complejidad y alternativas ofrecidas por la siguiente pregunta, difícilmente responderían utilizando un marco común de referencia y comprendiendo lo que se pide. "¿Cómo le va a usted y a su familia este año? ¿La pregunta se refiere a finanzas, felicidad marital, estado de salud o a qué?

   Se ha realizado un gran trabajo respecto a la redacción de reactivos. Si se siguen ciertos preceptos, esto ayuda al redactor a evitar ambigüedades. En primer lugar, deben evitarse las preguntas que contengan más de una idea a la que el entrevistado pueda reaccionar. Un reactivo como "¿considera usted que las metas educativas de la preparatoria moderna y los métodos de enseñanza utilizados para lograr estos objetivos son educativamente adecuados?" es una pregunta ambigua, debido a que al entrevistado se le cuestiona acerca de los objetivos educativos y de los métodos de enseñanza en la misma pregunta. En segundo lugar, se deben evitar términos y expresiones ambiguas. Es posible plantear la siguiente pregunta a un entrevistado: ¿piensa usted que los maestros de su escuela reciben un trato justo?" Se trata de un reactivo ambiguo debido a que "trato justo" puede referirse a diferentes tipos de trato. El término *justo* también puede significar "justicia", "equidad", "no demasiado bueno", "imparcial" y "objetivo". La pregunta requiere de un contexto claro, es decir, de un marco de referencia explícito. (Sin embargo, algunas preguntas ambiguas se utilizan deliberadamente para producir distintos marcos de referencia.)
4. *¿La pregunta es de tipo conducente?* Las preguntas conducentes sugieren respuestas y como tales, amenazan la validez. Si se le pregunta a alguien "¿ha leído usted sobre la situación de la escuela local?" se puede obtener un número desproporcionado de respuestas "sí", ya que la pregunta implicaría que es malo no haber leído sobre la situación de la escuela local.
5. *¿La pregunta demanda conocimiento e información que el entrevistado no posee?* Para contrarrestar la invalidez de una respuesta debida a falta de información, resulta sensato utilizar preguntas de filtro de información. Antes de preguntarle a una persona lo que piensa acerca de la UNESCO, primero debe preguntársele si sabe lo que significa y es la UNESCO. Existe otro método: se le explica al entrevistado brevemente lo que es la UNESCO y luego se le pregunta lo que piensa de ella.
6. *¿La pregunta demanda material personal o delicado que el entrevistado pueda negarse a proporcionar?* Se requiere de técnicas especiales para obtener información de naturaleza personal, delicada o polémica. Pregunte sobre los ingresos u otros asuntos personales más tarde en la entrevista, después de haber establecido el *rapport*. Si se pregunta respecto a algo que es desaprobado socialmente, primero debe mostrarse que algunas personas piensan en un sentido y que otras piensan en otro sentido. En

efecto, no debe provocarse que el entrevistado se desapruebe a sí mismo. Se le debe asegurar que todas las respuestas serán confidenciales.

7. *¿La pregunta está cargada de deseo de aceptación social?* La gente tiende a dar respuestas que son socialmente deseables, respuestas que indican o implican la aprobación de actos o cosas que son generalmente consideradas como "buenas". Se le puede preguntar a una persona sobre sus sentimientos hacia los niños. Se supone que todos deben amar a los niños. A menos que se sea cuidadoso, se obtendrá una respuesta estereotipada sobre los niños y el amor. También, si se le pregunta a una persona si vota, hay que tener cuidado, ya que se supone que todos deben votar. Si se le pregunta a un entrevistado sobre sus reacciones ante los grupos minoritarios, de nuevo se corre el riesgo de obtener respuestas inválidas. La mayor parte de las personas educadas, sin importar cuáles sean sus "verdaderas" actitudes, están conscientes de la desaprobación de los prejuicios. Entonces, una buena pregunta es aquella donde los entrevistados no son conducidos a expresar meros sentimientos socialmente deseables. Al mismo tiempo, tampoco se debe preguntar a los entrevistados de forma que se enfrenten con la necesidad de dar respuestas socialmente indeseables.

El valor de las entrevistas y de los inventarios de entrevistas

Cuando la entrevista se acompaña de un inventario adecuado de valor comprobado, constituye una herramienta de investigación potente e indispensable que produce datos que ninguna otra herramienta de investigación ofrece. Es adaptable, capaz de utilizarse con todo tipo de entrevistados en muchos tipos de investigaciones y única por su adecuación para hacer exploraciones profundas. ¿Pero equilibran sus fortalezas a sus debilidades? ¿Cuál es su valor en la investigación del comportamiento, en comparación con otros métodos de recolección de datos?

La herramienta más natural con la cual comparar a la entrevista es el llamado cuestionario. Como se señaló antes, "cuestionario" es un término que se emplea casi para cualquier clase de instrumento que incluye preguntas o reactivos a los que responde un individuo. Aunque el término se utiliza de manera intercambiable con "inventario", parece estar más asociado con instrumentos autoadministrados que poseen reactivos cerrados o de alternativa fija.

El *instrumento autoadministrado* posee ciertas ventajas. Siendo todos o la mayoría de sus reactivos de tipo cerrado, se alcanza mayor uniformidad de estímulo y, por lo tanto, mayor confiabilidad. A este respecto, tiene las ventajas de las escalas y pruebas escritas de tipo objetivo, si se elaboran y se prueban previamente de manera adecuada. Una segunda ventaja es que, si son anónimos y confidenciales, se alienta la honestidad y la franqueza. Este tipo de instrumento también se aplica a muchas personas de manera relativamente fácil. Una ventaja un tanto dudosa es que puede enviarse por correo a los entrevistados. Además, también es económica: su costo por lo general es una fracción del de las entrevistas.

Las desventajas de los instrumentos autoadministrados (cuando se envían por correo) parecen sobrepasar sus ventajas. La principal desventaja es un bajo porcentaje de recuperación. Una segunda desventaja es que quizá no sea tan uniforme como parece. La experiencia ha demostrado que con frecuencia la misma pregunta tiene diferente significado para distintas personas. Como se explicó antes, esto puede manejarse en la entrevista. Sin embargo, no es posible hacer algo para resolver dicha situación cuando se autoadministra el instrumento. En tercer lugar, si sólo se utilizan reactivos cerrados, el instrumento muestra las mismas debilidades de los reactivos cerrados analizadas con anterioridad. Por otro lado, si se utilizan reactivos abiertos, es posible que el entrevistado se niegue a escribir las

respuestas, lo cual reduce la muestra de respuestas adecuadas. Muchas personas no son capaces de expresarse adecuadamente a través de la escritura, y a muchos que sí pueden hacerlo, les disgusta.

Debido a estas desventajas, probablemente la entrevista sea superior al cuestionario autoadministrado. (Esta objeción, por supuesto, no incluye las escalas de personalidad y de actitud que están cuidadosamente elaboradas.) El mejor instrumento disponible para estudiar el comportamiento, las intenciones futuras, los sentimientos, las actitudes y las razones del comportamiento de las personas parece ser la entrevista estructurada, en combinación con un inventario de entrevista que incluya reactivos cerrados, abiertos y de escala. Por supuesto, la entrevista estructurada debe elaborarse y construirse con cuidado, así como aplicarse únicamente por entrevistadores hábiles. Sus principales desventajas son el costo en tiempo, energía y dinero, y el alto nivel de habilidad necesarios para su elaboración. Una vez que se superan sus desventajas, la entrevista estructurada se vuelve una poderosa herramienta.

El grupo focal y la entrevista de grupo: otro método de entrevista

Quizás este tema pertenezca a un capítulo previo cuando se expusieron los métodos cualitativos. Algunos investigadores equiparan al método del grupo focal con la investigación cualitativa (Calder, 1977). Algunos se han referido a este método como *entrevistas de grupo* (Wells, 1974). Basch (1987) reporta que este método fue expuesto por Bogardus en 1926, pero que sólo se utilizó ocasionalmente a partir de entonces hasta los años ochenta. Quienes utilizaban primordialmente el método del grupo focal, hasta hace poco, eran los investigadores de mercado y de negocios. Basch (1987) considera que el método del grupo focal es prometedor en áreas diferentes de la mercadotecnia. Él considera que podría ser una técnica de investigación para mejorar la investigación, práctica y teoría de la educación para la salud. El método proporciona una visión profunda de la gente. Sudman, Bradburn y Schwarz (1996) creen que la metodología del grupo focal sirve para determinar la manera en que los entrevistados producen y procesan información.

La técnica del grupo focal implica entrevistar a dos o más personas al mismo tiempo. El tamaño del grupo focal debe ser lo suficientemente grande para generar diversos puntos de vista, pero lo suficientemente pequeño para ser manejable. Krueger (1994) recomienda de siete a diez personas por grupo focal, lo cual permitirá a cada persona tener la oportunidad de participar en la discusión. Existe un moderador que conduce la discusión de forma abierta y libre. Este moderador o facilitador requiere estar bien entrenado. Es función del moderador hacer que la discusión no se aleje demasiado del tema de interés. El tema puede ser cualquiera. Las respuestas de los entrevistados no son solicitadas de forma activa. No se dan sugerencias directas. En investigación de mercado o de consumo el tema se referiría a un producto o servicio. En psicología el interés sería, por ejemplo, el lenguaje utilizado por hombres homosexuales afroamericanos (Mays, Cochran, Bellinger, Smith, Henley *et al.*, 1992). En el área de salud, un grupo focal se utilizaría para determinar los temores acerca de los cinturones de seguridad o las bolsas de aire. Una de las metas consiste en examinar las actitudes y el comportamiento de la gente. La otra meta es descubrir lo que cada participante piensa sobre el tema que se discute. Las opiniones y descripciones surgen de los entrevistados. El investigador espera ser capaz de descubrir, a través de las discusiones, los discernimientos importantes que después sirvan para resolver problemas. Calder (1977) afirma que el método del grupo focal es útil para descubrir infor-

mación que se utilice para diseñar un estudio cuantitativo de investigación. Algunos han utilizado el grupo focal como un medio para desarrollar cuestionarios. La investigación de grupo focal también ayuda a los investigadores a desarrollar constructos que empleen en estudios futuros. Calder lo llama "conocimiento precientífico".

Una de las ventajas de los grupos focales es su costo, pues cuesta muy poco organizarlos. Los mayores costos residirían en conseguir y pagar al moderador. Además, los participantes podrían recibir un pago simbólico por su tiempo. El grupo focal también se realiza de forma rápida. Se dispone de las ideas de los entrevistados rápidamente y se realiza una videofilmación de las sesiones para analizarlas después con mayor profundidad. Es muy bueno para generar hipótesis para posteriores investigaciones. En la investigación de mercado, el grupo focal permite al cliente (fabricante) que encargó el estudio, ser un participante activo en la participación grupal. Así, dicha persona es capaz de obtener la información de primera mano. Lo anterior es posible debido a que los grupos se organizan de un tamaño que sea manejable. La interacción entre los entrevistados puede generar intercambios estimulantes que resulten en información útil, que no se obtiene con otros métodos de investigación. Además, como se mencionó antes, los grupos focales son muy flexibles. Un moderador experto va dirigiendo, y aun permitiendo que ideas prometedoras fluyan.

Sin embargo, el grupo focal no es muy recomendable para producir información concreta. Una decisión no debe basarse únicamente en la información reunida con dicho método. Además, ha sido criticado por los investigadores cuantitativos como "no científico" e indigno de confianza. Las preguntas no son estandarizadas y pueden variar de un grupo a otro. Con el uso de grupos muy pequeños, los datos de los grupos focales sufren en su posibilidad de generalización. A diferencia de la investigación de encuesta estructurada, el grupo focal no implica mucho esfuerzo por asegurarse de que el grupo sea representativo. Como sucede en la dinámica de cualquier grupo, siempre habrá unos cuantos individuos que dominen la conversación. Entonces, el moderador necesita contar con suficiente experiencia para minimizar la situación sin cortar el flujo de comunicación. La entrevista de grupo focal requiere de mucha paciencia y habilidad. Berger (1991) ofrece algunas sugerencias útiles para el moderador. También bosqueja lo que debe contener un reporte sobre un grupo focal. Algunos participantes ven al grupo focal como una oportunidad para ventilar sus emociones. Por lo tanto, los temas sensibles no deben explorarse por medio de grupos focales. En la sección de sugerencias de estudio se presentan algunas muy buenas referencias sobre los grupos focales. Éstos constituyen un método cualitativo de investigación, y como tales son capaces de ofrecer información rica que no pueden explotar los métodos cuantitativos. Son muy adecuados para conocer lo que desean los clientes o lo que la gente piensa acerca de ciertas políticas y reglas. Los grupos enfocados han probado su eficacia en el estudio de organizaciones.

Algunos ejemplos de investigación de grupos focales

Audience Studies Incorporated (ASI) es una compañía de investigación de mercados que ha operado a las afueras de Hollywood, California, durante muchos años. Se invita a los consumidores a una demostración, en la que se presenta un programa y comerciales televisión. Por su participación se sortean premios tales como champú, crema dental, analgésicos, etcétera. Durante la demostración del programa y de los comerciales, los participantes utilizan un dispositivo electrónico de calificación para comunicar sus puntos de vista acerca de lo que están viendo. Las respuestas se graban. Los participantes también completan cuestionarios después de cada comercial o programa. A partir de este grupo se elige a varias personas para participar en grupos focales. Los fabricantes de productos generalmente comisionan a ASI para que conduzca estos grupos focales para obtener ideas sobre cómo funciona su producto en relación con la competencia. En varias ocasiones

participan representantes del fabricante en los grupos focales, para obtener información de manera directa. Por ejemplo, si un fabricante está pensando en desarrollar un nuevo producto, la información obtenida de los grupos focales brinda ideas sobre lo que debe llevar el producto en términos de manufactura y mercadotecnia.

Mays *et al.* (1992) utilizó un grupo focal que incluía hombres homosexuales afroamericanos. El tema de discusión era la conducta sexual y el VIH. Con este método, Mays *et al.* fueron capaces de recopilar el argot que emplean los varones afroamericanos homosexuales. Estos resultados son útiles para comparar a los hombres homosexuales afroamericanos con homosexuales americanos blancos, y también para construir cuestionarios diseñados para descubrir la conducta sexual de varones homosexuales afroamericanos. El conocimiento obtenido a partir de dicho estudio también serviría para educar a consejeros y profesionales de la salud que tratan con homosexuales afroamericanos. Mays y sus colaboradores (1992, p. 432) afirman lo siguiente:

> Al utilizar la terminología presentada aquí, en la conducción de investigación relacionada con el VIH, es importante recordar que los procesos lingüísticos y cognitivos están inmersos en un contexto. Al evaluar la conducta sexual de hombres homosexuales negros, el planteamiento de preguntas que incluyen su argot también debe originarse desde el marco de referencia de su experiencia.

Sussman, Burton, Dent, Stacy y Flay (1991) publicaron que se debe tener precaución con el uso de los grupos focales, pues consideran que tales grupos pueden inducir ciertos *efectos grupales* que sesguen las respuestas. Su estudio exploró el extenso procedimiento de los grupos focales, que incluye un cuestionario previo de grupo. El cuestionario tiene material que se cubrirá durante la sesión del grupo y puede afectar a los miembros del grupo al comprometerlos con una posición antes de que comience la discusión grupal. Estos investigadores consideran que la gente depende de las respuestas de otras personas, y de esta manera convergen en una norma colectiva. Es decir, algunos entrevistados tendrán juicios más extremos después de las discusiones grupales.

Uno de los efectos de las normas colectivas es el *efecto de polarización del grupo*. El involucrarse en un grupo puede sesgar a los participantes a responder de maneras más extremas. Específicamente Sussman *et al.* (1991) buscaron una polarización de actitudes (un efecto de sesgo por influencia del grupo). La discusión en el grupo focal se dirigió hacia cómo reclutar adolescentes que consuman tabaco para una clínica contra el tabaquismo. Se utilizaron 31 grupos focales; a cada uno se les administraron cuestionarios de pretest y de postest. Los datos obtenidos apoyaron la existencia de un efecto de polarización del grupo. Después de participar en un grupo focal, los entrevistados manifestaron una evaluación más alta de las estrategias de reclutamiento autogeneradas. También reportaron que si fuesen fumadores, dichas estrategias los inducirían a unirse al programa. El estudio demostró que los grupos focales podrían no generan nuevas estrategias. Sin embargo, sí parecen ser efectivos en inducir en los participantes una actitud más favorable hacia las soluciones autogeneradas de problemas.

RESUMEN DEL CAPÍTULO

1. La entrevista constituye el método más antiguo y universal para extraer grandes cantidades de información de la gente.
2. Los métodos de recolección de datos utilizados en una entrevista pueden ser clasificados de acuerdo a qué tan directos son en sus preguntas y planteamientos.

3. Las entrevistas requieren de mucho tiempo. Por lo tanto, la recolección de datos es costosa en términos de tiempo, esfuerzo y dinero.
4. Las entrevistas requieren de entrevistadores entrenados y de un cuestionario bien desarrollado.
5. La entrevista se utiliza para tres propósitos principales: como un dispositivo exploratorio para generar ideas e hipótesis, como el instrumento principal utilizado en un estudio y como complemento para otros métodos y/o como instrumento de seguimiento.
6. La entrevista es una situación interpersonal cara a cara. Existen dos tipos: estructurada o no estructurada.
7. Un tipo de entrevista no estructurada es la entrevista de grupo o grupos focales.
8. Se buscan tres tipos de información en los inventarios de entrevista: de identificación, de tipo censal (sociológica) y del problema.
9. Los tipos de reactivos utilizados en un inventario de entrevista son: reactivos de alternativa fija, reactivos abiertos y reactivos de escala.
10. Existen siete criterios para la redacción de reactivos y preguntas en el inventario.
11. El método del grupo focal es una entrevista no estructurada que utiliza un número pequeño de participantes. Es de bajo costo y de rápida realización.
12. La investigación de grupos focales es de tipo cualitativo.
13. Los grupos focales tienen un problema de generalización.

SUGERENCIAS DE ESTUDIO

1. A continuación se incluyen varias referencias valiosas sobre la entrevista y unas cuantas sobre el inventario de entrevista.

Obras clásicas

Cannell, C. y Kahn, R. (1968). Interviewing, en G. Lindzey y E. Aronson (eds.), *The handbook of social psychology*, vol. II (2a. ed.). Reading, MA: Addison-Wesley, 526-595.

Survey Research Center. (1976). *Interviewer's manual* (ed. rev.). Ann Arbor, Michigan: Institute for Social Research, University of Michigan. [Una excelente guía sobre los aspectos prácticos de la entrevista.]

Warwick, D. y Lininger, C. (1975). *The sample survey: Theory and practice.* Nueva York: McGraw-Hill.

Trabajos más recientes

Beed, T. W. y Stimson, R. J. (1985). *Survey Interviewing: Theory and techniques.* Nueva York: Routledge, Chapman, and Hall.

Bowden, J. C. (1995). *An investigator's guide to interviewing and interrogation.* Orlando, Florida: Bowden.

Knale, S. (1996). *Interviews: An introduction to qualitative research interviewing.* Thousand Oaks, California: Sage.

Lukas, S. (1993). *Where to start and what to ask: The assessment handbook.* Nueva York: Norton.

Mollica, R. F. y Caspi-Yavin, Y. (1991). Measuring torture and torture-related symptoms. *Psychological Assessment, 3*, 581-587. [Explica por qué los instrumentos y técnicas actuales de entrevista son inadecuados cuando se usan para entrevistar personas que han sido torturadas.]

Myers, J. (1996). *Interviewing young children about body touch and handling.* Chicago, Illinois: University of Chicago Press.

2. Por fortuna existen buenos inventarios de entrevista en abundancia. El lector debe estudiar uno o dos de ellos con cuidado. Los inventarios sugeridos a continuación están bien construidos y son muy interesantes. Note que los inventarios publicados casi siempre incluyen explicaciones metodológicas extensas. El estudiante aprenderá mucho sobre la construcción de escalas de entrevista con el estudio de este material.

Campbell, A., Converse, P. y Rodgers, W. (1976). *The quality of American life.* Nueva York: Russell Sage Foundation, app. B. [Un inventario grande con muchos reactivos de escala y cuidadosas instrucciones para el entrevistador. También es un estudio muy importante.]

Free, L. y Cantril, H. (1967). *The political beliefs of Americans.* New Brunswick, Nueva Jersey: Rutgers University Press, app. B. [Presenta buenas preguntas, sondeos y reactivos de alternativa fija.]

Glock, C. y Stark, R. (1966). *Christian beliefs and anti-Semitism.* Nueva York: Harper & Row. [Al final del libro se presenta el inventario completo, principalmente con reactivos de alternativa fija.]

3. Los ejemplos dados en el punto 2 son todos de investigación de encuesta, el campo de investigación donde el arte y la técnica de la entrevista se desarrolló y utilizó en primera instancia. Sin embargo, las entrevistas son y han sido utilizadas en lo que puede denominarse estudios "normales", es decir, estudios cuyo único o principal interés es encontrar relaciones entre variables. El estudio de Burt (1980) sobre las actitudes hacia la violación es un buen ejemplo; a continuación se mencionan otros ejemplos.

Estudios "normales"

Beckman, L. J. y Mays, V. M. (1985). Educating community gatekeepers about alcohol abuse in women: Changing attitudes, knowledge and referral practices. *Journal of Drug Education, 15,* 289-309. [Se utilizó una entrevista telefónica para evaluar el efecto de dos talleres sobre el conocimiento, actitudes y prácticas de referencia hacia las mujeres que sufren de abuso de alcohol.]

Campbell, A. y Schuman, H. (1968). *Racial Attitudes in fifteen cities.* Ann Arbor, Michigan: Institute for Social Research, University of Michigan. [Una combinación de "encuesta" y preguntas actitudinales enfocadas a la comprensión de las actitudes raciales y su cambio.]

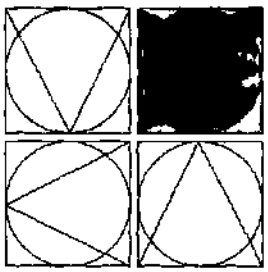
Doob, A. y MacDonald, G. (1979). Television viewing and fear of victimization: Is the relationship causal? *Journal of Personality and Social Psychology, 37,* 170-179. [Se realizó una entrevista de puerta en puerta para determinar si la gente que ve más la televisión tiene más temores que la gente que la ve menos. Se realizó una comparación entre aquellos que viven en un área de baja criminalidad y quienes viven en un área de alta criminalidad.]

Gersch, I. S. y Nolan, A. (1994). Exclusion: What the children think. *Educational Psychology in Practice, 10,* 35-45. [Se diseñó, aplicó y analizó un inventario de entrevista para medir las actitudes y experiencias de los niños hacia la escuela. Dicho instrumento se utilizó para evaluar a estudiantes que fueron excluidos.]

Jones, S. L. (1996). The association between objective and subjective care-giver burden. *Archives of Psychiatric Nursing, 10*, 77-84. [En tres momentos de recolección de datos se utilizaron entrevistas telefónicas para encontrar asociaciones entre las cargas objetivas y subjetivas de cuidadores.]

4. Los siguientes son libros y artículos que tratan con la teoría y práctica de grupos focales en ciencias sociales y del comportamiento. Los grupos focales con frecuencia sirven como generadores de ideas y para comprobar hipótesis en un ambiente informal. Consiga uno de ellos y lea los capítulos sobre metodología.

Berger, A. A. (1991). *Media research techniques.* Newbury Park, California: Sage.
Greenbaum, T. L. (1993). *The handbook for focus group research.* Nueva York: Lexington Books.
Morgan, D. L. (1988). *Focus groups as qualitative research.* Newbury Park, California: Sage.
Templeton, J. F. (1994). *The focus group: A strategic guide to organizing, conducting, and analyzing the focus group interview.* Chicago, Illinois: Probus Publishing.
Vaughn, S. (1996). *Focus group interviews in education and psychology.* Thousand Oaks, California: Sage.



CAPÍTULO 30

PRUEBAS Y ESCALAS OBJETIVAS

El método de observación y recolección de datos más utilizado en las ciencias del comportamiento es el de las pruebas y escalas. La gran cantidad de tiempo que pasan los investigadores en la construcción o búsqueda de medidas de variables es tiempo bien invertido, ya que la adecuada medición de las variables de investigación constituye uno de los aspectos nodales del trabajo científico sobre el comportamiento. En general, se ha puesto muy poca atención a la medición de las variables de estudios de investigación. ¿Qué utilidad pueden tener los intrigantes e importantes problemas de investigación, el sofisticado diseño de investigación y el intrincado análisis estadístico, si las variables de los estudios de investigación se miden pobremente? Por fortuna se ha progresado mucho en la comprensión de la teoría de la medición psicológica y educativa, así como en el mejoramiento de la práctica de la medición. En este capítulo se examina algo de la tecnología que subyace a los procedimientos objetivos de medición.

Objetividad y métodos objetivos de observación

La objetividad, una característica central y esencial de la metodología científica, es fácil de definir, aunque evidentemente difícil de entender. También es polémica. La objetividad es el acuerdo entre jueces expertos respecto a lo que se observa. Los métodos objetivos de observación son aquellos donde, cualquiera que siga las reglas prescritas, asignará los mismos valores numéricos a objetos y conjuntos de objetos como lo haría cualquier otra persona. Un procedimiento objetivo es aquel en el que el acuerdo entre los observadores se encuentra en su nivel máximo. En términos de varianza, la varianza de los observadores se encuentra en su nivel mínimo, lo cual quiere decir que la varianza de juicio, es decir, la varianza debida a las diferencias entre jueces en la asignación de valores numéricos a objetos, se aproxima a cero. Es posible encontrar una extensa discusión sobre la objetividad en Kerlinger (1979, pp. 9-13 y 262-264). La importancia de la comprensión de la objetividad en la ciencia no puede sobrestimarse. Es especialmente importante comprender que la objetividad científica es metodológica y tiene poco o nada que ver con la objetividad como una supuesta característica de los científicos. El hecho de que un científico como persona sea o no sea objetivo no es el punto importante. Lo importante es que la objetividad científica es inherente a los procedimientos metodológicos, caracterizados por el acuerdo entre jueces expertos —y nada más—.

Todos los métodos de observación son inferenciales: se realizan inferencias respecto a las propiedades de los miembros de conjuntos con base en los valores numéricos asignados a tales miembros, por medio de entrevistas, pruebas, escalas y observaciones directas del comportamiento. Los métodos difieren respecto a si son directos o indirectos, en el grado en que las inferencias son realizadas a partir de las observaciones en bruto. Las inferencias realizadas por medio de métodos objetivos de observación, por lo general, son muy largas, a pesar de que aparentan ser directas. La mayoría de dichos métodos permiten un alto grado de acuerdo entre los observadores, ya que los participantes registran marcas en papel y tales marcas están restringidas a dos o más opciones de alternativas dadas por el observador. A partir de las marcas en el papel, el observador infiere las características de los individuos y conjuntos de individuos que hacen las marcas. En un tipo de métodos objetivos, el observador (o juez) hace las marcas en papel, observa al objeto o los objetos de medición y elige entre las alternativas dadas. En tal caso, también se realizan inferencias sobre las propiedades del objeto o los objetos observados, a partir de las marcas en papel. La principal diferencia reside en quién hace las marcas.

Debe reconocerse que todos los métodos de observación poseen cierta objetividad. No existe una clara dicotomía, en otras palabras, entre los llamados métodos objetivos y otros métodos de observación. Más bien existe una diferencia en el grado de objetividad. Nuevamente, si se piensa en los grados de objetividad como grados de acuerdo entre observadores, desaparece la ambigüedad y confusión que con frecuencia se asocia con el problema.

Entonces existe el acuerdo de que lo que aquí se llama métodos objetivos de observación y medición no poseen el monopolio de objetividad ni de inferencia, sino que son más objetivos y no menos inferenciales que cualquier otro método de observación y medición. Los métodos que se expondrán en el presente capítulo por ningún motivo abarcan todos los métodos posibles, pues el tema es grande y muy variado. Se consideran únicamente como medidas de variables, vistas y evaluadas de la misma forma que todas las demás medidas de variables.

Pruebas y escalas: definiciones

Una *prueba* es un procedimiento sistemático en el que se presenta a los individuos un conjunto de estímulos construidos a los cuales responden. Las respuestas permiten que quien realiza la prueba asigne a los examinados valores numéricos o conjuntos de valores numéricos, a partir de los cuales se pueden realizar inferencias sobre si el examinado posee aquello que se supone que la prueba está midiendo. Esta definición dice un poco más que la que afirma que la prueba es un instrumento de medición.

Una *escala* es un conjunto de símbolos o valores numéricos, construida de tal manera que los símbolos o valores numéricos puedan ser asignados por una regla a los individuos (o a sus comportamientos) a quienes se aplica la escala, y donde la asignación indica si el individuo posee lo que se supone que mide la escala. Al igual que una prueba, una escala es un instrumento de medición. De hecho, con excepción del significado excesivo asociado con la prueba, se observa que prueba y escala se definen de manera similar. Sin embargo, estrictamente hablando, la escala se utiliza de dos maneras: para indicar un instrumento de medición y para indicar los valores numéricos sistematizados del instrumento de medición. Se utiliza en ambos sentidos sin demasiada preocupación por las distinciones. No obstante, se debe recordar que: *las pruebas son escalas, pero las escalas no necesariamente son pruebas*. Es así porque las escalas, por lo general, no tienen el significado de competencia ni de éxito o fracaso que tienen las pruebas.

Tipos de medidas objetivas

La mayor parte de los cientos, quizás miles, de pruebas y escalas pueden dividirse en las siguientes clases: pruebas de inteligencia y aptitud, pruebas de rendimiento, medidas de personalidad, escalas de actitud y valores, y medidas objetivas diversas. A continuación se discutirá cada uno de estos tipos de medida, desde un punto de vista de investigación.

Pruebas de inteligencia y aptitud

En la investigación psicológica y educativa con frecuencia se necesita de una medida de inteligencia o aptitud, ya sea como variable independiente o como variable dependiente. Para evaluar los efectos de programas educativos de uno u otro tipo sobre el rendimiento académico, por ejemplo, generalmente es necesario controlar la inteligencia, de tal manera que las diferencias encontradas entre los grupos de tratamiento experimental no puedan atribuirse a diferencias en la inteligencia, más que a los tratamientos. Existe un gran número de buenas pruebas de inteligencia que utilizan los investigadores, quizás sea el caso de una de las llamadas pruebas de recopilación (ómnibus). (Una prueba de *recopilación* es aquella que tiene diferentes tipos de reactivos: verbales, numéricos, espaciales y otros, en un instrumento.) Dichas pruebas por lo general son altamente verbales y se correlacionan de forma sustancial con el rendimiento académico. Los manuales de Buros (1998) son guías útiles para dichas pruebas. Anastasi (1988) ofrece una lista clasificada de pruebas representativas en su libro, y divide cada prueba en clasificaciones separadas, tales como de inteligencia, de personalidad, etcétera.

Una *aptitud* es la habilidad potencial para el logro. Las pruebas de aptitud se utilizan principalmente para orientación y consejería. También se emplean en investigación, particularmente como variables control. Una *variable control* es aquella cuyo efecto sobre una variable dependiente quizá requiera anularse. Por ejemplo, al estudiar el efecto de un programa correctivo de lectura sobre el rendimiento en lectura, puede ser necesario atribuir la aptitud verbal a las posibles diferencias de grupo en habilidad verbal. De forma

similar, tal vez sea necesario controlar otras variables de influencia potencial: habilidades numéricas y espaciales, por ejemplo. Las pruebas de aptitud resultan útiles en tales casos.

Pruebas de rendimiento

Las *pruebas de rendimiento* miden la eficiencia, maestría y comprensión presentes, en áreas generales y específicas del conocimiento. En su mayoría, constituyen mediciones de la eficacia de la instrucción y el aprendizaje y son, por supuesto, enormemente importantes en educación y en la investigación educativa. De hecho, como se mencionó antes, con frecuencia el rendimiento es la variable dependiente en investigaciones que incluyen métodos de instrucción.

Las pruebas de rendimiento a menudo se clasifican de varias formas. Para los propósitos de este libro, se dividen, primero, en pruebas estandarizadas y en pruebas de construcción especial. Las *pruebas estandarizadas* son grupos de pruebas ya publicadas, que se basan en un contenido educativo general común a un gran número de sistemas educativos. Son los productos de un alto grado de competencia y habilidad profesional en la redacción de pruebas y, como tales, por lo general son bastante confiables y casi siempre válidas. Están dotadas con minuciosas tablas de normas (promedios) que pueden utilizarse con propósitos comparativos. Las *pruebas de construcción especial* son generalmente pruebas realizadas *ex profeso* por maestros para medir logros más específicos y limitados. Por supuesto, también pueden ser elaboradas por investigadores educativos para medir áreas limitadas de rendimiento.

En segundo lugar, las pruebas estandarizadas de rendimiento pueden, a su vez, clasificarse en pruebas generales y especiales. Las *pruebas generales* son baterías de pruebas que miden las áreas más importantes del rendimiento académico: uso del lenguaje, vocabulario, lectura, aritmética y estudios sociales. Las pruebas de rendimiento especial, como su nombre lo indica, son pruebas de materias individuales tales como historia, ciencia e inglés.

Los investigadores rara vez eligen las pruebas de rendimiento debido a que los sistemas escolares son quienes las seleccionan. Sin embargo, cuando a los investigadores se les da la oportunidad de elegir, deben evaluar cuidadosamente el tipo de prueba de rendimiento que requiere su problema de investigación. Suponga que la variable de investigación en un estudio es "el rendimiento en la comprensión de conceptos". Muchas, quizás la mayoría de las pruebas utilizadas en las escuelas, no son adecuadas para medir dicha variable. En tales casos, los investigadores pueden elegir una prueba especialmente diseñada para medir la comprensión de conceptos, o ellos mismos diseñar la prueba. La construcción de una prueba de rendimiento es un trabajo enorme, aunque no es posible comentar aquí los detalles. Se refiere al estudiante a pruebas especializadas. Por desgracia, existen pocos libros sobre la construcción de pruebas y escalas para propósitos de investigación. La mayoría de los libros y otros textos sobre medición se enfocan, casi siempre, en la construcción y el uso de instrumentos con propósitos de aplicación. Sin embargo, los investigadores que necesiten construir medidas de rendimiento de cualquier tipo encontrarán una excelente guía en los trabajos de Adkins (1974); Cangelosi (1990); Gronlund (1988); Haladyna (1997); Hopkins (1989), y Osterlind (1989). Los investigadores que necesiten construir escalas de actitud encontrarán en el libro de Edwards (1957) un documento invaluable. Dawes (1972) cubre algunos métodos que Edwards no cubrió.

Medidas de personalidad

La medición de rasgos de personalidad constituye el problema más complejo de la medición psicológica. La razón es simple: la personalidad humana es extremadamente compleja. Para propósitos de medición, la personalidad puede considerarse como la organización particular de los rasgos del individuo. Un *rasgo* es una característica de un individuo, revelada

a través de comportamientos recurrentes en situaciones diferentes. Se dice que un individuo es compulsivo o que posee el rasgo de compulsividad debido a que se observa que esa persona es notoriamente pulcra tanto en su vestimenta como en su lenguaje, siempre es puntual, desea que todo esté muy ordenado, y le disgustan y evita las irregularidades.

El principal problema en la medición de la personalidad es la validez. Medir la validez de los rasgos de personalidad requiere que se conozca qué son tales rasgos, la forma en que interactúan y cambian, y cómo se relacionan entre sí, lo cual constituye un requisito muy grande e incluso severo. La cuestión no es, como les gusta señalar a los críticos ingenuos, que la personalidad no pueda medirse debido a que es demasiado esquiva, demasiado compleja, demasiado existencial, o que los esfuerzos de medición no hayan tenido éxito, sino más bien que se ha logrado cierto grado de éxito en una tarea tan difícil. No obstante, el problema de la validez es considerable.

Existen dos modelos generales para la construcción y validación de medidas de personalidad: el *método a priori* y el *método teórico* o *de constructo*. En el *método a priori*, los reactivos se construyen para reflejar la dimensión de personalidad que se mide. Puesto que el introvertido con frecuencia es una persona que se aísla, quizá se redacten reactivos sobre la preferencia por estar solo. Por ejemplo, esto podría incluir reactivos que indiquen la preferencia por evitar fiestas para medir la introversión. Como la persona ansiosa estará nerviosa y desorganizada cuando se encuentre bajo estrés, se podrían redactar reactivos que sugieran estas condiciones, para medir la ansiedad. Entonces, en el método *a priori* el redactor de la escala reúne o redacta reactivos que midan, de forma ostensible, rasgos de personalidad.

Éste es esencialmente el modelo de los primeros redactores de pruebas de personalidad. Aunque no existe nada inherentemente malo con el método —de hecho, tendrá que utilizarse, especialmente en las primeras etapas de la construcción de pruebas y escalas— los resultados llegan a ser confusos. Los reactivos no siempre miden lo que se cree que están midiendo. En algunas ocasiones incluso se descubre que un reactivo que se pensaba que medía, por ejemplo, responsabilidad social, en realidad mide una tendencia a estar de acuerdo con afirmaciones socialmente deseables. Por tal razón, el método a priori, utilizado de forma única, resulta insuficiente. Por lo anterior, las pruebas de personalidad generalmente carecen de validez de contenido.

El método de validación que se utiliza a menudo con las escalas de personalidad *a priori* es el *método de grupo conocido*. Para validar una escala de responsabilidad social, es posible buscar un grupo de individuos con un alto nivel conocido de responsabilidad social, y otro con un bajo nivel conocido de responsabilidad social. Si la escala logra diferenciar con éxito a ambos grupos, entonces se considera que tiene validez.

La medida de personalidad *a priori* y otras medidas continuarán empleándose en la investigación del comportamiento. Sin embargo, debe evitarse su uso ingenuo y a ciegas. Debe verificarse su validez de constructo y su validez relacionada con el criterio, especialmente por medio del análisis factorial y otras formas empíricas. Las medidas de personalidad, al igual que otras medidas, se han utilizado con frecuencia tan sólo porque quienes las usan piensan que miden aquello que afirman que están midiendo.

El *método teórico* o *de constructo* de la construcción de medidas de personalidad enfatiza las relaciones de la variable medida con otras variables, las relaciones surgidas de la teoría que subyace a la investigación. Aunque la construcción de escalas debe ser siempre, en cierto grado, *a priori*, mientras mayor sea el número de medidas de personalidad sujetas a las pruebas de validez de constructo, mayor será la confianza en su fidelidad. No es suficiente aceptar simplemente la validez de una escala de personalidad ni aun aceptar su validez debido a que ha logrado diferenciar con éxito, por ejemplo, a artistas de científicos, a maestros de no maestros, a personas normales de personas neuróticas. A final de cuentas,

debe establecerse su validez de constructo, es decir, su uso exitoso en la predicción de una gran variedad de relaciones teóricas.

Escalas de actitud

Las actitudes, aunque se tratan de forma separada aquí y en la mayor parte de los análisis de libros de texto, en realidad son parte integral de la personalidad. Los teóricos modernos también consideran la inteligencia y la aptitud como partes de la personalidad. Sin embargo, la medición de la personalidad trata principalmente de rasgos. Un *rasgo*, como se mencionó en la sección anterior, es una característica relativamente perdurable del individuo, a responder de cierta manera en todas las situaciones. Si alguien es dominante, exhibirá comportamiento dominante en la mayoría de las situaciones. Si alguien es ansioso, presentará una conducta ansiosa en la mayor parte de sus actividades. Por otro lado, una *actitud* es una predisposición organizada a pensar, sentir, percibir y comportarse hacia un referente u objeto cognitivo. Se trata de una estructura perdurable de creencias que predispone al individuo a comportarse de manera selectiva hacia los referentes de actitud.

Un *referente* es una categoría, una clase o un conjunto de fenómenos: objetos físicos, eventos, conductas e inclusive constructos (véase Brown, 1958). La gente tiene actitudes hacia muchas cosas diferentes: grupos étnicos, instituciones, religión, aspectos y prácticas educativas, la Suprema Corte, derechos civiles, propiedad privada, etcétera. Dicho en otras palabras, se tienen actitudes hacia algo "de afuera". Un rasgo tiene una referencia subjetiva; una actitud tiene una referencia objetiva. Alguien que tiene una actitud hostil hacia los extranjeros quizá sea hostil únicamente con los extranjeros; pero alguien que tiene un rasgo de hostilidad es hostil hacia todas las personas (al menos potencialmente).

Existen tres tipos principales de escalas de actitud: escalas de puntuación sumada, escalas de intervalos aparentemente iguales y escalas acumulativas (o Guttman). Una escala de puntuación sumada (un tipo de las cuales se denomina escala tipo Likert) es un conjunto de reactivos de actitud, todos los cuales son considerados con un "valor de actitud" aproximadamente igual, y donde cada uno de los participantes responde con grados de acuerdo o desacuerdo (intensidad). Las puntuaciones de los reactivos de dicha escala se suman, o se suman y promedian, para producir una puntuación de actitud del individuo. Como en todas las escalas de actitud, el propósito de la escala de puntuación sumada es ubicar a un individuo en algún punto de un continuo del nivel de acuerdo de la actitud en cuestión.

Es importante hacer notar dos o tres características de las escalas de puntuación sumada, pues muchas escalas comparten estas características. Primero, U, el universo de reactivos, se considera un conjunto de reactivos con igual "valor de actitud", como se indicó en la definición anterior. Ello significa que no existe una escala de reactivos como tal; un reactivo es igual a cualquier otro reactivo respecto a su valor de actitud. Se "clasifica" a los individuos que responden los reactivos. La clasificación surge de las sumas (o promedios) de las respuestas de los individuos. Cualquier subconjunto de U es teóricamente igual que cualquier otro subconjunto de U: se ordenaría a un conjunto de individuos por rango de la misma manera si se utiliza U_2 o U_1.

En segundo lugar, las escalas de puntuación sumada permiten la expresión de la intensidad de la actitud. Los participantes pueden estar simplemente de acuerdo o estar fuertemente de acuerdo. Esto conlleva ventajas, así como desventajas. La principal ventaja es que resulta una mayor varianza. Cuando existen cinco o siete categorías posibles de respuesta, resulta obvio que la varianza de la respuesta debe ser mayor que con sólo dos o tres categorías (por ejemplo, de acuerdo, en desacuerdo y sin opinión). Por desgracia, la varianza de las escalas de puntuación sumada con frecuencia parece contener varianza debida a la fijeza de respuesta. Los individuos tienen tendencias diferentes a usar ciertos tipos

de respuestas: respuestas extremas, respuestas neutrales, respuestas en acuerdo y respuestas en desacuerdo. Dicha varianza de respuesta confunde la varianza de la actitud (y el rasgo de personalidad). Las diferencias individuales producidas por las escalas de actitud de puntuaciones sumadas (y medidas de rasgos calificadas de manera similar) han mostrado deberse, en parte, a la fijeza de respuesta y a otras extrañas fuentes de varianza similares. La literatura sobre la fijeza de respuesta es muy amplia y no puede citarse en detalle. La exposición de Nunnally (1978) está bien equilibrada. Mientras que la fijeza de respuesta puede considerarse como una ligera amenaza para la validez de la medición, su importancia ha sido sobrestimada y la evidencia disponible no justifica las fuertes afirmaciones negativas hechas por los partidarios de la fijeza de respuesta. En otras palabras, mientras se debe estar consciente de las posibilidades y amenazas, no hay que paralizarse por el incremento del riesgo (véase Rorer, 1965, para un análisis más profundo).

A continuación se presentan dos reactivos de puntuación sumada de una escala construida por Burt (1980, p. 222) para el estudio de actitudes hacia la violación. Tales reactivos se desarrollaron para medir el estereotipo del rol sexual. Se utilizó una escala de 7 puntos que va desde un fuerte acuerdo (7) hasta un fuerte desacuerdo (1). Los valores entre paréntesis (y los valores intermedios) se asignan a las respuestas indicadas.

Hay algo malo con la mujer que no desea casarse y formar una familia.
Una mujer debe ser virgen cuando se casa.

Las escalas de intervalos aparentemente iguales de Thurstone se construyen a partir de principios diferentes. Mientras que el producto final, un conjunto de reactivos de actitud, puede utilizarse para el mismo propósito de asignación de puntuaciones individuales de actitud, las escalas de intervalos aparentemente iguales también logran el importante propósito de escalar los reactivos de actitud. A cada reactivo se le asigna un valor de escala que indica la fortaleza de actitud de una respuesta de acuerdo para el reactivo. El universo de reactivos se considera un conjunto ordenado; es decir, los reactivos difieren en su valor de escala. El procedimiento de clasificación encuentra estos valores de escala. Además, los reactivos de la escala final a utilizarse son tan selectos que los intervalos entre ellos son iguales, lo cual representa una importante y deseable característica psicométrica.

Los siguientes reactivos de intervalos aparentemente iguales, con los valores de escala de los reactivos, provienen de la escala de Actitud hacia la Iglesia, de Thurstone y Chave (1929, pp. 61-63, 78):

Considero que la Iglesia es la mayor institución en Estados Unidos hoy. (Valor de escala: 0.2)
Creo en la religión, pero en raras ocasiones voy a la iglesia. (Valor de escala: 5.4)
Pienso que la Iglesia es un obstáculo para la religión, ya que aun depende de lo sobrenatural, la superstición y el mito. (Valor de escala: 9.6)

En la escala de Thurstone y Chave, a menor valor de escala del reactivo, más positiva sería la actitud hacia la Iglesia. El primer y tercer reactivos son respectivamente el más bajo y el más alto en la escala. El segundo reactivo, por supuesto, tiene un valor intermedio. La escala total contenía 45 reactivos con valores de escala a través de todo el continuo. Sin embargo, por lo general, las escalas de intervalos aparentemente iguales contienen bastante menos reactivos.

El tercer tipo de escala, la *escala acumulativa* o *Guttman*, consta de un conjunto relativamente pequeño de reactivos homogéneos que son unidimensionales (o que supuestamente lo son). Una escala unidimensional mide una variable y sólo una. La escala obtiene su nombre de la relación acumulativa entre los reactivos y las puntuaciones totales de los

individuos. Por ejemplo, a cuatro niños se les plantean tres preguntas aritméticas: (a) 28/7 = ?, (b) 8 × 4 = ? y (c) 12 + 9 = ? El niño 1, quien resuelve (a) correctamente, tiene muchas posibilidades de resolver (b) y (c) también correctamente. El niño 2, quien resuelve (a) de forma incorrecta pero (b) de forma correcta, también tiene muchas posibilidades de resolver (c) correctamente. El niño 3, quien resuelve correctamente sólo (c), tiene muy pocas probabilidades de resolver (a) y (b) de forma correcta. La situación se sintetiza de la siguiente manera (la tabla incluye la puntuación del cuarto niño, quien no resuelve ninguna pregunta de forma correcta):

	(a)	(b)	(c)	Puntuación total
Niño 1	1	1	1	3
Niño 2	0	1	1	2
Niño 3	0	0	1	1
Niño 4	0	0	0	0

(1 = Correcto; 0 = Incorrecto)

Note la relación entre el patrón de respuestas a los reactivos y las puntuaciones totales. Si se conoce la puntuación total de un niño, se puede predecir su patrón, si la escala es acumulativa, solamente si el conocimiento de las respuestas correctas de los reactivos más difíciles predice las respuestas de los reactivos más fáciles. Observe también que se escalan ambos reactivos y las personas.

De manera similar, es posible plantear a la gente varias preguntas acerca de un objeto actitudinal. Si bajo análisis los patrones de respuesta se ordenan entre sí de la forma indicada arriba (o por lo menos de manera muy similar), entonces se dice que las preguntas o reactivos son unidimensionales. De esta manera, las personas pueden ser ordenadas de acuerdo con sus respuestas en la escala (véase Edwards, 1957, capítulo 7, para una mayor discusión sobre las escalas acumulativas unidimensionales).

Resulta obvio que estos tres métodos de construcción de escalas de actitud son muy diferentes. Considere que métodos similares o iguales pueden utilizarse con otros tipos de escalas de personalidad u otras escalas. La escala de puntuación sumada se concentra en los participantes y sus ubicaciones dentro de la escala. La escala de intervalos aparentemente iguales se concentra en los reactivos y sus ubicaciones dentro de la escala. De manera interesante, ambos tipos de escalas producen casi los mismos resultados en lo que se refiere a la confiabilidad y a la ubicación de los individuos en órdenes de rango actitudinales. Las escalas acumulativas se concentran en la posibilidad de escalar de los conjuntos de reactivos y en la posición de los individuos en la escala.

De los tres tipos de escalas, la de puntuaciones sumadas parece ser el más útil para la investigación del comportamiento, pues es más fácil de desarrollar y, como se indicó antes, produce casi los mismos resultados que la escala de intervalos aparentemente iguales, de construcción más laboriosa. Utilizadas con cuidado y con el conocimiento de sus debilidades, las escalas de puntuaciones sumadas se adaptan a muchas de las necesidades de los investigadores del comportamiento. Las escalas acumulativas parecen ser de menor utilidad y menos aplicables de manera general. Si se utiliza un objeto cognitivo de corte claro, una escala acumulativa breve y bien construida genera medidas confiables de un número de variables psicológicas: tolerancia, conformismo, identificación grupal, aceptación de la autoridad, permisividad, etcétera. También debe señalarse que el método puede mejorarse y alterarse de varias formas. Dawe (1972) y Edwards (1957) describen la forma de construir y evaluar escalas acumulativas, así como escalas de puntuación sumada y escalas de intervalos aparentemente iguales.

Escalas de valores

Los *valores* son preferencias de tipo cultural hacia objetos, ideas, gente, instituciones y comportamientos (Kluckhohn, 1951, pp. 388-433). Mientras que las actitudes son organizaciones de creencias sobre cosas "de afuera", es decir, predisposiciones a comportarse hacia los objetos o referentes de actitudes, los valores expresan preferencias por formas de conducta y estados de existencia (Rokeach, 1968). Términos como *igualdad*, *religión*, *libre empresa*, *derechos civiles* y *obediencia* expresan valores. Dicho de manera sencilla, los valores expresan lo "bueno", lo "malo", los "debería" y los "debiera" del comportamiento humano. Los valores colocan las ideas, las cosas y los comportamientos en un continuo de aprobación-desaprobación. Implican opciones entre cursos de acción y de pensamiento.

Con el propósito de brindar al lector una idea de los valores, se presentan tres reactivos. Se puede pedir a los individuos que expresen su aprobación o desaprobación sobre el primero y segundo reactivos, quizás en forma de puntuación sumada, y que elijan una de las tres alternativas del tercer reactivo.

> Por el propio bien y por el bien de la sociedad, una persona debe ser controlada por la tradición y la autoridad.
> Ahora más que nunca debemos fortalecer la familia, el estabilizador natural de la sociedad.
> ¿Cuál de los siguientes aspectos es el más importante para desarrollar una vida plena: la educación, el logro o la amistad?

Por desgracia, los valores han recibido escasa atención científica, aun cuando éstos y las actitudes conforman gran parte de la producción verbal de las personas y son, probablemente, influyentes determinantes del comportamiento. Por lo tanto, la medición de valores padece desatención. Sin embargo, los valores sociales y educativos tal vez se conviertan en el centro de mucho más trabajo teórico y empírico en el futuro, ya que los científicos sociales se han vuelto cada vez más conscientes de que los valores constituyen influencias importantes en el comportamiento individual y de grupo (véase Dukes, 1955; Haddock y Zanna, 1998; Hendrick, Hendrick y Dicke, 1998: Hogan, 1973; Lubinski, Schmidt y Benbow, 1996; Pittel y Mendelsohn, 1966; Robinson, 1996). Una fuente de escalas de valores es Levitin (1969). Un ensayo muy sugestivo y valioso que apareció hace 45 años es el trabajo realizado por Kluckhohn (1951). Otro ensayo sobre la medición de valores que aún es importante es el de Thurstone (1959).

Tipos de escalas y reactivos objetivos

Dos tipos generales de reactivos que se usan con frecuencia son aquellos en que las respuestas son independientes y aquellos en que no son independientes. *Independencia* aquí significa que la respuesta de una persona a un reactivo no está relacionada con su respuesta a otro reactivo. Todos los reactivos de verdadero-falso, *sí-no*, *de acuerdo-en desacuerdo* y de tipo Likert pertenecen al tipo independiente. El sujeto responde cada reactivo libremente, con un rango de dos o más respuestas posibles, de las cuales puede elegir sólo una. Por otro lado, los reactivos no independientes obligan al sujeto a elegir un reactivo o alternativa que excluye la elección de otros reactivos o alternativas. Tales formas de escalas y reactivos se denominan de elección forzada. El sujeto se enfrenta con dos o más reactivos o subreactivos y se le pide que elija uno o más de ellos de acuerdo con algún criterio, e incluso criterios.

Dos ejemplos simples mostrarán la diferencia entre reactivos independientes y no independientes. Primero, es posible dar al sujeto un conjunto de instrucciones que permitan

independencia de respuesta; en segundo lugar, se le puede indicar un conjunto de instrucciones contrastante, con opciones más limitadas (no independientes):

Ejemplos

> Indique junto a cada una de las siguientes afirmaciones qué tanto las aprueba, utilizando una escala del 1 al 5, donde el 1 significa "No lo apruebo en lo absoluto" y el 5 significa "Lo apruebo muchísimo".
>
> A continuación se dan 40 pares de afirmaciones. De cada par escoja la que apruebe más. Márquela con una palomita (✓).

Las ventajas de los reactivos independientes son la economía y aplicabilidad de la mayoría de los análisis estadísticos a sus respuestas. Además, cuando se responde cada reactivo se obtiene un máximo de información, donde cada reactivo contribuye a la varianza. También se requiere de menos tiempo para aplicar las escalas independientes, aunque quizá sufran del sesgo provocado por la fijeza de respuesta. Los individuos pueden dar respuestas similares o iguales para cada reactivo: pueden apoyarlos todos con entusiasmo o con indiferencia dependiendo de su predilección particular de respuesta. La varianza importante de una variable llega, entonces, a confundirse por la fijeza de respuesta.

El tipo de escala de elección forzada evita, por lo menos en cierto grado, el sesgo de respuesta. Sin embargo, al mismo tiempo, sufre por la falta de independencia, y por su costo excesivo y alta complejidad. No obstante, existen algunos investigadores, como Comrey (1970), que han construido una escala del sesgo de respuesta dentro de la prueba de personalidad. Las escalas de elección forzada en ocasiones también agotan la tolerancia y la paciencia del sujeto, lo cual se traduce en menor cooperación. Aun así, muchos expertos consideran que los instrumentos de elección forzada son prometedores para la medición educativa y psicológica. Mientras que otros expertos se muestran escépticos.

Entonces, las escalas y los reactivos se dividen en tres tipos: *de acuerdo-en desacuerdo* (o aprobación-desaprobación o *verdadero-falso*, y similares), de orden de rango y de elección forzada. Cada uno de ellos se analiza brevemente. En la literatura pueden encontrarse explicaciones más detalladas (véase Edwards, 1957; Guilford, 1954).

Reactivos de acuerdo-desacuerdo

Existen tres formas generales de reactivos de acuerdo y desacuerdo:

1. Aquellos que permiten una de dos respuestas posibles.
2. Aquellos que permiten una de tres o más respuestas posibles.
3. Aquellos que permiten más de una elección de tres o más respuestas posibles.

Las dos primeras formas proporcionan alternativas como "de acuerdo-desacuerdo"; "sí-no"; "lo apruebo-sin opinión-lo desapruebo"; "lo apruebo mucho–lo apruebo-lo desapruebo mucho"; "1, 2, 3, 4, 5". Los participantes eligen una de las respuestas proporcionadas para reportar sus reacciones a los reactivos. Al hacerlo, dan reportes sobre sí mismos o indican sus reacciones a los reactivos. La mayor parte de las escalas de personalidad y de actitud utilizan dichos reactivos. Si una persona está construyendo un instrumento que utilice este método, es muy importante la manera en que se redacte la escala. Por ejemplo, suponga que se elaboró la siguiente escala de Likert de 5 puntos:

1 = En desacuerdo
2 = En ligero desacuerdo
3 = Neutral

4 = En ligero acuerdo
5 = De acuerdo

El problema aquí se centra en la manera en que el lector interpreta el significado de cada punto de la escala. Un sujeto puede elegir un "2" y otro puede elegir un "4". Ambos pudieron haber hecho la misma interpretación. Si alguien está en ligero acuerdo, entonces esa persona quizá también esté en ligero desacuerdo. De ahí surge la confusión.

El tercer tipo de escala de este tipo presenta un número de reactivos: se dan instrucciones a los participantes para indicar aquellos reactivos que los describen, reactivos con los que están de acuerdo o simplemente reactivos que ellos eligen. El listado de adjetivos es un buen ejemplo. Se le presenta al sujeto una lista de adjetivos, donde algunos indican rasgos deseables como analítico, generoso y considerado; y donde otros indican rasgos indeseables como cruel, egoísta y vulgar. Se les pide que marquen aquellos adjetivos que los caracterizan. (Por supuesto, este tipo de instrumento también sirve para caracterizar a otras personas.) Quizás una forma mejor sería una lista con todos los adjetivos positivos de escalas de valores conocidas, donde se les pide a los participantes que seleccionen un número específico de sus propias características personales. La escala de intervalos aparentemente iguales y su sistema de respuesta donde se marcan los reactivos de actitud con los que se está de acuerdo es, por supuesto, la misma idea. La idea es útil, especialmente con el desarrollo de escalas factoriales, de métodos de escalación y el creciente uso de los métodos de elección.

La calificación de los reactivos de acuerdo-en desacuerdo llega a ser problemática debido a que no todos los reactivos, o los componentes, reciben respuestas. (Con una escala de puntuación sumada o una escala de puntuación ordinaria, los participantes generalmente responden a todos los reactivos.) Sin embargo, en general, se pueden utilizar sistemas simples de asignación de valores numéricos a las diversas opciones. Por ejemplo, de acuerdo-en desacuerdo puede ser 1 y 0; sí-no puede ser 1, 0, –1 o, evitando los signos negativos: 2, 1, 0. A las respuestas de los reactivos de puntuación sumada descritos anteriormente simplemente se les asigna del 1 al 5 o del 1 al 7.

El principal aspecto que los investigadores deben tener en mente es que el sistema de puntuación debe producir datos interpretables, congruentes con el sistema de puntuación. Si se utilizan puntuaciones de 1, 0, –1, los datos deben ser capaces de proveer una interpretación escalada; es decir, 1 es "alto" o "mucho", –1 es "bajo" o "poco" y 0 está en medio. Un sistema de 1, 0 puede significar alto y bajo o simplemente presencia o ausencia de un atributo. Tal sistema puede ser útil y poderoso, como se vio anteriormente cuando se estudiaron variables como sexo, raza, clase social, etcétera. En síntesis, los datos producidos por sistemas de puntuación deben tener significados claramente interpretables en cierto sentido cuantitativo. Se refiere al lector al análisis de Ghiselli (1964, pp. 44-49) sobre el significado de las puntuaciones. No obstante, algunos expertos han criticado el uso de 0-1 o de los sistemas binarios de puntuación. Durante el desarrollo de sus escalas de personalidad, Comrey descubrió que los reactivos que utilizan un esquema de respuesta binaria están sujetos a problemas y distorsiones que no necesariamente se obtendrían si la escala tuviera 3 puntos o más. Los resultados del estudio de Comrey sobre las escalas se resumen en Comrey (1978) y Comrey y Lee (1992).

Se han desarrollado varios sistemas para ponderar reactivos; aunque la evidencia indica que las puntuaciones ponderadas y no ponderadas dan en gran parte los mismos resultados. Los estudiantes parecen encontrar esto difícil de creer. (Note que se habla acerca de la ponderación de las respuestas a los reactivos.) Aunque el asunto no está completamente establecido, existe fuerte evidencia de que en pruebas y medidas con el suficiente número de reactivos —como 20 o más— la ponderación diferencial de reactivos no genera mucha diferencia en los resultados finales. Tampoco la ponderación diferencial de respuestas pro-

duce mucha diferencia (véase Guilford, 1954; Nunnally, 1978). Tampoco se produce ninguna diferencia, en términos de varianza, si se transforma las ponderaciones de las puntuaciones de manera lineal. Se puede hacer que los participantes utilicen un sistema, +1, 0, –1 y, por supuesto, utilizar las puntuaciones en un análisis. Sin embargo, se puede añadir una constante de 1 a cada puntuación, produciendo 2, 1, 0. Las puntuaciones transformadas son más fáciles de trabajar, ya que no tienen signos negativos.

Reactivos y escalas de orden de rango

El segundo grupo de tipos de escalas y reactivos es ordinal o de orden de rangos, que es una forma simple y muy útil de escala o reactivo. Una escala completa puede ordenarse por rangos; es decir, se le pide a los participantes que ordenen todos los reactivos de acuerdo a algún criterio específico. Por ejemplo, si se desea comparar los valores educativos de administradores, maestros y padres, se les presenta un número de reactivos que presuntamente midan valores educativos a los miembros de cada grupo con las instrucciones de que los ordenen de acuerdo con sus preferencias.

En su estudio sobre actitudes hacia la liberación femenina, Taleporos (1977) desarrolló una escala de orden de rango de problemas sociales. Se les pidió a los participantes que ordenaran los siguientes problemas sociales: *adicción a las drogas, contaminación ambiental, discriminación racial, discriminación sexual, crimen violento* y *asistencia social*. Taleporos esperaba que los dos grupos que estaba estudiando ordenaran los temas sociales de manera similar, con excepción del tema de la *discriminación sexual*. Se sostuvo su hipótesis. Su estudio representó un uso productivo de la escalación de orden de rangos.

Las escalas de orden de rangos tienen tres ventajas analíticamente convenientes:

1. Las escalas de los individuos pueden interrelacionar y analizar fácilmente. Los órdenes de rangos compuestos de los grupos de individuos también se correlacionan fácilmente.
2. Los valores de escala de un conjunto de estímulos pueden calcularse utilizando uno de los métodos de orden de rango de escalación (véase Guilford, 1954).
3. Las escalas escapan parcialmente de la fijeza de respuesta y de la tendencia a mostrarse de acuerdo con los reactivos socialmente deseables.

Reactivos y escalas de elección forzada

La esencia de un método de elección forzada es que el sujeto debe elegir entre alternativas que en apariencia se perciben casi igualmente favorables (o desfavorables). Estrictamente hablando, el método no es nuevo. Las escalas de comparaciones de pares y de orden de rangos son métodos de elección forzada. Lo que es diferente acerca del método de elección forzada, como tal, es que se determinan los valores de discriminación y preferencia de los reactivos, y se aparean aquellos reactivos que son aproximadamente iguales en ambos. Así se controlan en cierta medida, la fijeza de respuesta y la "deseabilidad del reactivo". (La *deseabilidad del reactivo* significa que un reactivo puede ser elegido sobre otro simplemente porque expresa una cuestión deseable reconocida comúnmente. Si a un hombre se le pregunta si es descuidado o eficiente, tiende a decir que es eficiente, aunque sea descuidado.)

El método de las comparaciones apareadas (o comparaciones de pares) posee un largo y respetable pasado psicométrico. Sin embargo, se ha utilizado principalmente con el propósito de determinar valores de escala (Guilford, 1954). Aquí las comparaciones de pares se consideran como un método de medición. La esencia del método es que conjuntos de pares de estímulos, o reactivos de diferentes valores en un solo continuo o en dos continuos o factores diferentes, se presentan al sujeto con las instrucciones de que elija un miembro de cada par con base en algún criterio establecido. El criterio podría ser: el que

mejor caracterice al sujeto o el que el sujeto prefiera. Los reactivos de los pares pueden ser palabras solas, enunciados e inclusive párrafos. Por ejemplo, Edwards, en su inventario de preferencia personal (Personal Preference Schedule), apareó de manera efectiva afirmaciones que expresan distintas necesidades. Un reactivo que mide la necesidad de autonomía, por ejemplo, está apareado con otro reactivo que mide la necesidad de cambio. Se le pide al sujeto que elija uno de esos reactivos. Se supone que la persona elegirá el reactivo que se ajuste a sus necesidades. Una característica única de la escala es que los valores del deseo de aceptación social de los miembros apareados se determinaron de forma empírica y los pares se relacionaron de acuerdo con esto. El instrumento produce perfiles de puntuaciones de necesidad para cada individuo.

De alguna manera los dos tipos de técnicas de comparaciones de pares, 1) la determinación de valores de escala de estímulos y 2) la medición directa de variables, constituyen los métodos psicométricos más satisfactorios. Son simples y económicos debido a que solamente existen dos alternativas. Además, se puede obtener una gran cantidad de información con una cantidad limitada de material. Por ejemplo, si un investigador tiene únicamente 10 reactivos, cinco de la variable *A* y cinco de la variable *B*, se puede construir una escala de 5×5 o 25 reactivos, ya que cada reactivo *A* puede aparearse sistemáticamente con cada reactivo *B*. (La puntuación es simple: asignar un "1" a *A* o a *B* en cada reactivo, dependiendo de la alternativa que el sujeto elija.) Más importante aún, los reactivos de comparación de pares obligan a los participantes a elegir. Aunque esto puede molestar a algunos participantes, especialmente si consideran que ningún reactivo representa lo que elegirían (es decir, elegir entre cobarde y débil para categorizarse a sí mismo), es en realidad una actividad humana acostumbrada. Debemos elegir cada día de nuestras vidas. Inclusive se puede argumentar que los reactivos de acuerdo-en desacuerdo son artificiales y que los reactivos de elección son "naturales". En un estudio sobre el concepto del interés social de Adler (valorar cosas diferentes al yo), Crandall (1980) utilizó comparaciones de pares para desarrollar su escala de interés social (Social Interest Scale). Los jueces calificaron 90 rasgos respecto a su relevancia para el interés social. Se utilizaron 48 pares, donde un miembro de cada par tenía relevancia para el interés social y el otro miembro no la tenía. Entonces, después de la realización de un análisis de reactivos para determinar cuáles eran los reactivos más discriminantes, se desarrolló una escala de 15 reactivos. Por desgracia, Crandall no reporta la forma de la escala. Sin embargo, la idea es buena: utilizó la fortaleza de las comparaciones de pares para encontrar buenos reactivos para una escala final.

Los reactivos de elección forzada con más de dos partes pueden asumir un número de formas con tres, cuatro o cinco partes, las cuales son homogéneas o heterogéneas respecto a lo favorable y a lo no favorable. Se analiza e ilustra sólo uno de estos tipos para demostrar los principios que subyacen a dichos reactivos. Por medio de un análisis factorial, un procedimiento conocido como la *técnica de los incidentes críticos* o algún otro método, se reúnen y seleccionan los reactivos. Por lo común se descubre que algunos reactivos discriminan entre grupos de criterio y que otros no lo hacen. Ambos tipos de reactivos —llámense discriminantes e irrelevantes— se incluyen en cada conjunto de reactivos. Además, se determinan los valores de preferencia para cada reactivo.

Un reactivo típico de elección forzada es una *tétrada*. Una forma útil de tétrada consiste en dos pares de reactivos, un par con un alto valor de preferencia y el otro par con un bajo valor de preferencia, donde un miembro de cada par es discriminativo (válido) y el otro miembro del par es irrelevante (no válido). Un esquema de dicho reactivo de elección forzada es:

alta preferencia-discriminante
alta preferencia-irrelevante

baja preferencia-discriminante
baja preferencia-irrelevante

Se dirige al sujeto para que elija el reactivo de la tétrada que más prefiera, o que constituya la mejor descripción de sí mismo (o de alguien más), etcétera. También se dirige a esta persona para que seleccione el reactivo menos preferido o menos descriptivo de sí mismo.

La idea básica detrás de este reactivo más bien complejo es, como se indicó antes, que se controla la fijeza de respuesta y el deseo de aceptación social. El sujeto no puede decir, al menos teóricamente, cuáles son los reactivos discriminantes y cuáles los irrelevantes; tampoco se pueden elegir los reactivos con base en los valores de preferencia. Así, se contrarresta la tendencia a evaluarse a sí mismo (o a otros) demasiado alto o demasiado bajo y, por lo tanto, la validez presuntamente se incrementa (Guilford, 1954).

Un reactivo de elección forzada de un tipo hasta cierto punto diferente, construido por el primer autor de este libro con fines ilustrativos del uso de reactivos de investigación real, es:

consciente
agradable
respondiente
sensible

Uno de los reactivos (sensible) es un reactivo *A* y otro (consciente) es un reactivo *B*. (*A* y *B* se refieren a factores adjetivados.) Los otros reactivos son presuntamente irrelevantes. Se puede pedir a los participantes que elijan uno o dos reactivos que son muy importantes que un maestro posea.

Los métodos de elección forzada parecen ser muy promisorios. Aun así, existen dificultades técnicas y psicológicas, entre las cuales la más importante parece ser la falta de independencia de los reactivos, la quizás demasiado compleja naturaleza de algunos reactivos y la resistencia de los participantes ante las opciones difíciles. Se refiere al lector a los estudios de Guilford (1954) o de Bock y Jones (1968) sobre el tema: éstos son de autoridad, objetivas y breves; y también a las revisiones de Scott (1968) y Zavala (1965). (Para la revisión de referencias más recientes sobre reactivos y escalas de elección forzada véase Borg, 1988; Bownas y Bernardin, 1991; Closs, 1978; Deaton, Glasnapp y Poggio, 1980; Hyman y Sharp, 1983; May y Forsyth, 1980; Presser y Schuman, 1980; Ray, 1990; y Stanley, Wandzilak, Ansorge y Potter, 1987.)

Medidas ipsativas y normativas

Una distinción que se ha vuelto importante y que generalmente es mal entendida, en investigación y medición, es aquella que existe entre las medidas normativas e ipsativas. Las *medidas normativas* son el tipo común de medidas obtenidas con pruebas y escalas; pueden variar de manera independiente, es decir, se ven relativamente poco afectadas por otras medidas y, para su interpretación, se refieren a la media de las medidas de un grupo, siendo que los conjuntos de medidas de individuos poseen medias y desviaciones estándar diferentes. Las *medidas ipsativas*, por otro lado, se ven afectadas de manera sistemática por otras medidas y, para su interpretación, se refieren a la misma media, siendo que cada conjunto de medidas del individuo posee la misma media y desviación estándar. Para terminar con esta más bien opaca palabrería, sólo piense en un conjunto de rangos, del 1 al 5, donde el 1 indica "el primero", "el más alto" o "el más"; y el 5 indica "el último", "el más bajo" y "el menos", con el 2, 3 y 4 señalando posiciones intermedias. Sin importar quién utilice estos rangos, la suma y la media de los rangos es siempre la misma, 15 y 3, y la desviación estándar es siempre la misma, 1.414. Los rangos, entonces, son medidas ipsativas.

Si los valores 1, 2, 3, 4 y 5 estuvieran disponibles para calificar, por ejemplo, cinco objetos, y fueran cuatro personas las que calificaran los cinco objetos, se obtendría algo como lo siguiente:

	Personas			
	1	2	2	3
	2	2	1	2
Objetos	3	4	5	3
	4	3	5	3
	5	5	4	2
Sumas:	15	16	17	13
Medias:	3.0	3.2	3.4	2.6

Note que las sumas y las medias (y las desviaciones estándar también) son diferentes. Éstas son medidas normativas. Teóricamente con las medidas normativas no existen restricciones en el valor que el individuo *A* le puede asignar al objeto *C*—con excepción, por supuesto, de los números del 1 al 5—.

No obstante, con las medidas ipsativas, el procedimiento —en este caso de orden de rangos— ha creado restricciones sistemáticas. Cada individuo debe utilizar 1, 2, 3, 4 y 5 tan sólo una vez, y todos deben ser utilizados, lo cual indica que cuando cinco objetos están siendo ordenados por rango y se asigna uno, por ejemplo, rango 1, sólo quedan cuatro rangos por asignar. Después de que se asigna el 2 al siguiente objeto, sólo quedan tres, etcétera, hasta el último objeto, al que debe asignársele el 5. Un razonamiento similar se aplica a otro tipo de procedimientos y medidas ipsativos: comparaciones de pares, tétradas y pentadas de elección forzada o metodología Q.

La limitación importante en los procedimientos ipsativos es que, estrictamente hablando, no se pueden aplicar los estadísticos usuales, puesto que éstos dependen de los supuestos que los procedimientos ipsativos violan sistemáticamente. Además, el procedimiento ipsativo genera correlaciones negativas espurias entre reactivos. En un instrumento de comparaciones de pares, por ejemplo, la selección de un miembro de un par automáticamente excluye la selección del otro miembro. Ello significa una falta de independencia y una correlación negativa entre reactivos, en función del procedimiento instrumental. Sin embargo, la mayor parte de las pruebas estadísticas se basan en el supuesto de independencia de los elementos que entran en las fórmulas estadísticas. Además, el análisis de correlaciones, tal como el análisis factorial, puede distorsionarse seriamente por las correlaciones negativas. Por desgracia, estas limitaciones no se han comprendido o se han subestimado por los investigadores que, por ejemplo, han tratado los datos ipsativos de forma normativa (Hicks, 1970). Se invita al lector a demostrar el comportamiento de las escalas ipsativas estableciendo una pequeña matriz de números ipsativos, generados de forma hipotética por medio de las respuestas en una escala de comparaciones de pares. Utilice números 1 y 0 y calcule las r entre reactivos para los individuos.

Elección y construcción de medidas objetivas

Una de las tareas más difíciles para el investigador del comportamiento, cuando se enfrenta con la necesidad de medir variables, es encontrar el camino a través de un gran número de medidas ya existentes. Si existe una buena medida para una variable en particular, parece

tener poco sentido construir una medida nueva. De cualquier manera, la pregunta es: ¿existe una buena medida? La respuesta a esta pregunta quizá requiera de una gran búsqueda y estudio. El investigador debe saber, primero, qué tipo de variable se va a medir. Se ha tratado de ofrecer una guía dentro de la estructura recién proporcionada. Se debe saber claramente si la variable es una aptitud, rendimiento, personalidad, actitud o algún otro tipo de variable. El segundo paso es consultar uno o dos libros de texto que analicen medidas y pruebas psicológicas. Después, se deben consultar las bien conocidas guías de Buros. Aunque Buros ofrece una excelente guía sobre pruebas publicadas, muchas buenas medidas no se han publicado comercialmente. Por lo tanto, debe buscarse en la literatura de aparición periódica. A pesar de que muchas escalas no están disponibles de manera comercial, se pueden reproducir (con permiso) y utilizarse con propósitos de investigación. Otras fuentes valiosas son Andrulis (1977); Comrey, Backer y Glaser (1973); Fischer y Corcoran (1994); Goldman, Saunders y Busch (1996); Keyser y Sweetland (1987) y Taulbee (1983).

Fuentes valiosas de información sobre pruebas y escalas son las revistas *Psychological Bulletin*, *Journal of Psychoeducational Assessment*, *Applied Psychological Measurement*, *Educational and Psychological Measurement*, *Journal of Educational Measurement*, *Psychological Assessment* y *Journal of Experimental Education*.

Tal vez un investigador encuentre que no existe una medida que mida el atributo deseado. O, si existe una medida, quizá sea insatisfactoria para los propósitos. Por consiguiente, el investigador debe construir una nueva medida o instrumento, o abandonar la variable. La construcción de pruebas y escalas objetivas constituye una tarea larga y ardua. No existen atajos. Un instrumento pobremente construido llega a ocasionar más daño que beneficio, ya que puede conducir al investigador a conclusiones erróneas. Entonces, el investigador que debe construir un nuevo instrumento tiene que seguir ciertos procedimientos conocidos y guiarse por criterios psicométricos aceptados.

Se ha logrado un enorme progreso en la medición objetiva de la inteligencia, las aptitudes, el rendimiento, la personalidad y las actitudes. Sin embargo, las opiniones están divididas, en ocasiones de forma muy marcada, sobre el valor de la medición objetiva. El avance más impresionante se ha logrado en la medición objetiva de la inteligencia, las aptitudes y el rendimiento. Los avances en la medición de la personalidad y las actitudes no han sido tan impresionantes. El problema es, por supuesto, la validez, especialmente la validez de las medidas de personalidad.

Dos o tres desarrollos recientes son muy alentadores. Uno es la creciente comprensión de la complejidad que implica la medición de cualquier variable de personalidad y de actitud. El segundo lo constituyen los avances técnicos para llevarla a cabo. Otro desarrollo muy relacionado consiste en el empleo del análisis factorial como ayuda en la identificación de variables y como guía en la construcción de medidas. Un tercer desarrollo (que se estudió en un capítulo anterior) es el creciente conocimiento, comprensión y maestría del problema de validez en sí mismo, y en especial la comprensión de que la validez y la teoría psicológica están interrelacionados.

RESUMEN DEL CAPÍTULO

1. La prueba o escala es el método más utilizado en las ciencias del comportamiento para la recolección de datos.
2. Una meta consiste en desarrollar y utilizar pruebas que sean objetivas. Sin embargo, la objetividad no es fácil de comprender.
3. La objetividad científica no depende de las características del científico.

4. La objetividad científica incluye el acuerdo entre jueces expertos. Los métodos de observación y la recolección de datos poseen diferentes grados de objetividad.
5. Una prueba es un procedimiento sistemático para determinar el comportamiento de individuos.
6. Una escala es un conjunto de símbolos o valores numéricos construidos de tal manera que tales símbolos o valores numéricos puedan asignarse a individuos utilizando alguna regla.
7. Las pruebas de aptitud miden el potencial de logro de la persona. Se usan principalmente para orientación y consejería.
8. Las pruebas de rendimiento miden la eficiencia, maestría y comprensión presentes, de áreas generales y específicas del conocimiento. Las pruebas elaboradas por maestros son consideradas como pruebas de rendimiento.
9. La medición de los rasgos de personalidad constituye el problema más complejo de la medición psicológica. La personalidad es muy compleja respecto a problemas de validez.
10. Existen dos métodos para la construcción y validación de medidas de personalidad: el método *a priori* y el método *de constructo*.
11. Las escalas de actitud miden la predisposición de un individuo a pensar, sentir, percibir y comportarse hacia otra persona, idea u objeto.
12. Existen tres tipos de escalas de actitud: la escala de puntuaciones sumadas, las escalas de intervalos aparentemente iguales y las escalas acumulativas o Guttman.
13. La escala de puntuaciones sumadas es la escala utilizada con mayor frecuencia en las ciencias del comportamiento.
14. Las escalas de valores miden la preferencia expresada por una persona hacia formas de conducta. Incluyen religión y libre empresa.
15. Existen dos tipos de escalas objetivas: independientes y no independientes.
16. En las escalas objetivas independientes la respuesta de una persona a un reactivo no se relaciona con su respuesta a otro reactivo. En los reactivos no independientes una respuesta a un reactivo podría conducir al sujeto a preguntas más profundas.
17. Las escalas y los reactivos pueden dividirse en tres tipos: *de acuerdo-en desacuerdo*, de orden de rangos y de elección forzada.
18. Las medidas normativas no se ven afectadas por otras medidas. Sin embargo, las medidas ipsativas sí son afectadas por otras medidas.
19. El investigador debe dedicar tiempo para determinar si ya existe una prueba para el estudio. Se cuenta con un número de fuentes publicadas y no publicadas de pruebas. Sólo se debe crear una prueba nueva si no existe una para los propósitos del investigador.

SUGERENCIAS DE ESTUDIO

1. Las siguientes referencias ayudarán a los estudiantes a encontrar su camino en el largo y difícil, pero importante, camino de las pruebas y escalas objetivas, especialmente en educación.

Adkins, D. (1974). *Test construction: Development and interpretation of achievement tests* (2a. ed.) Columbus, Ohio: Charles E. Merrill. [Un libro invaluable para estudiantes e investigadores.]

Bloom, B. (ed.). (1976). *Taxonomy of educational objectives. The classification of educational goals: Handbook 1, cognitive domain*. Nueva York: David McKay. [Este libro básí-

co e inusual intenta establecer un fundamento para la medición cognitiva, por medio de la clasificación de los objetivos educativos y por la presentación de muchos preceptos y ejemplos. Las páginas 201-207, que bosquejan el libro, son útiles para quienes construyen pruebas y para los investigadores educativos.]

Impara, J. C. y Plake, B. S. (eds.). (1998). *Buros 13th mental measurements yearbook.* Lincoln, Nebraska: Buros Institute. [Descripciones y revisiones de pruebas publicadas y medidas de todos tipos. Véase también ediciones anteriores.]

Mehrens, W. y Ebel, R. (eds.). (1967). *Principles of educational and psychological measurement.* Chicago: Rand McNally. [Una valiosa colección de muchas de las contribuciones clásicas a la medición y a la teoría y práctica de las pruebas.]

2. Para comprender la lógica y la elaboración de instrumentos psicológicos de medición, resulta útil estudiar las explicaciones relativamente completas sobre cómo se desarrollan. Las siguientes referencias escritas describen el desarrollo de interesantes e importantes reactivos e instrumentos de medición.

Allport, G., Vernon, P. y Lindzey, G. (1951). *Study of values. Manual of directions* (ed. rev.). Boston: Houghton Mifflin.

Comrey, A. L. (1961). Factored homogeneous item dimensions in personality research. *Educational and Psychological Measurement, 21,* 417-431.

Comrey, A. L. y Lee, H. B. (1992). *A first course in factor analysis* (2a. ed.). Hillsdale, Nueva Jersey: Lawrence Erlbaum.

Edwards, A. (1953). *Personal preference schedule, manual.* Nueva York: Psychological Corp. [Mide necesidades en formato de elección forzada (comparaciones de pares).]

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology, 140.* [Monografía original de Likert donde describe su técnica; se trata de una importante guía en la medición de actitudes.]

Thurstone, L. y Chave, E. (1929). *The measurement of attitude.* Chicago: University of Chicago Press. [Este clásico describe la construcción de la escala de intervalos aparentemente iguales, para medir actitudes hacia la Iglesia.]

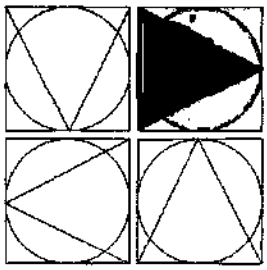
Woodmansee, J. y Cook, S. (1967). Dimensions of verbal racial attitudes: Their identification and measurement. *Journal of Personality and Social Psychology, 7.* 240-250. [Probablemente la mejor medida para actitudes hacia los negros. El inventario aparece en el volumen de Robinson, Rusk y Head, citado en las sugerencias de estudio 3.]

3. Las siguientes son tres útiles antologías de escalas de actitud, valores y otras escalas. Su utilidad reside no sólo en las muchas escalas que contienen, sino también en las críticas perspicaces que se enfocan en la confiabilidad, la validez y otras características de las escalas.

Robinson, J., Rusk, J. y Head, K. (1968). *Measures of political attitudes.* Ann Arbor: Institute for Social Research, University of Michigan.

Robinson, J. y Shaver, P. (1969). *Measures of social psychological attitudes.* Ann Arbor: Institute for Social Research, University of Michigan.

Shaw, M. y Wright, J. (1967). *Scales for the measurement of attitudes.* Nueva York: McGraw-Hill.



CAPÍTULO 31

OBSERVACIONES DEL COMPORTAMIENTO Y SOCIOMETRÍA

Todos observan los actos de otros. Se observa a otras personas y se les escucha hablar. Se infiere lo que los otros quieren decir cuando expresan algo; y se infieren las características, motivaciones, sentimientos e intenciones de otros con base en estas observaciones. Se dice

"ella es una juez perspicaz de la gente", queriendo decir que sus observaciones sobre el comportamiento son agudas y que se considera que sus inferencias sobre lo que está detrás del comportamiento son válidas. Sin embargo, este tipo de observaciones diarias de la mayoría de la gente resultan insatisfactorias para la ciencia. Los científicos sociales también deben observar el comportamiento humano; pero no se satisfacen con observaciones no controladas. Ellos buscan observaciones confiables y objetivas a partir de las cuales puedan realizar inferencias válidas. Tratan la observación del comportamiento como parte de un procedimiento de medición: asignan valores numéricos a objetos de acuerdo con reglas, en este caso de acuerdo con actos o secuencias de actos de comportamiento humano.

A pesar de que esto puede parecer simple y directo, evidentemente no lo es: existe mucha controversia y debate respecto a la observación y a los métodos de observación. Los críticos del punto de vista de que las observaciones del comportamiento deben ser controladas rigurosamente —punto de vista adoptado en este capítulo y en el resto del libro— afirman que es demasiado estrecho y artificial. Los críticos dicen que en lugar de eso, las observaciones deben ser naturales: los observadores deben estar inmersos en situaciones realistas y naturales presentes, y deben observar el comportamiento tal y como ocurre de manera natural, por así decirlo. Sin embargo, como se verá, la observación del comportamiento es extremadamente compleja y difícil.

Existen básicamente dos formas de observación: se puede observar a la gente hacer y decir cosas, y se puede preguntar a la gente sobre sus propias acciones y sobre el comportamiento de otros. Las principales maneras para obtener información son experimentando algo de forma directa o pidiéndole a alguien que informe lo que sucedió. En el presente capítulo se incluyen principalmente los eventos que se ven y se escuchan y la observación del comportamiento, así como la solución de problemas científicos que surgen a partir de dichas observaciones. También se examina de manera breve un método para evaluar las interacciones e interrelaciones de miembros de grupos: la sociometría, la cual es una forma especial y valiosa de observación. Los miembros de grupos se observan entre sí y registran las reacciones entre ellos, de tal manera que los investigadores evalúen el estado sociométrico de los grupos.

Problemas en la observación del comportamiento

El observador

El principal problema de la observación del comportamiento reside en el observador. Una de las dificultades de las entrevistas es que el entrevistador forma parte del instrumento de medición. Dicho problema casi no existe en las pruebas y escalas objetivas. En la observación del comportamiento, el observador es tanto una fortaleza, como una debilidad cruciales. Esto se debe a que el observador debe asimilar la información derivada de las observaciones, para luego realizar inferencias acerca de los constructos. El observador mira cierto comportamiento —por ejemplo, un niño que golpea a otro niño— y debe procesar, de alguna manera, tal observación y hacer una inferencia de que el comportamiento es una manifestación del constructo "agresión" o "comportamiento agresivo", o incluso "hostilidad". La fortaleza y debilidad del procedimiento es el poder de inferencia del observador. Si no fuera por la inferencia, una máquina observadora sería mejor que un observador humano. La fortaleza radica en que el observador puede relacionar el comportamiento observado con el constructo o variables de un estudio al unir el comportamiento con el constructo. Una de las dificultades recurrentes de la medición consiste en lograr cerrar el abismo que existe entre el comportamiento y el constructo.

La debilidad básica del observador consiste en que se pueden realizar inferencias incorrectas a partir de las observaciones. Considere dos casos extremos. Suponga, por un lado, que un observador, que se muestra muy hostil ante la educación escolar religiosa, observa las clases en una escuela religiosa. Queda claro que los prejuicios de esta persona pueden invalidar las observaciones. El observador tal vez califique a un maestro adaptable como inflexible debido a la existencia de un prejuicio o a la percepción de que la enseñanza en escuelas religiosas es inflexible. O tal vez ese mismo observador juzgue el comportamiento realmente estimulante de un maestro de una escuela religiosa como insulso. Por otro lado, suponga que un observador pueda ser completamente objetivo y no sabe nada sobre la educación pública o religiosa. En cierto sentido cualquier observación que realice no estará sesgada; pero será inadecuada. La observación del comportamiento humano requiere de un conocimiento competente sobre dicho comportamiento y aun del significado del comportamiento.

Sin embargo, existe otro problema: el observador puede afectar los objetos de observación en tanto que forma parte de la situación de observación. Sin embargo, en realidad y por fortuna, éste no constituye un problema severo. De hecho, representa un problema para el novato, quien parece creer que la gente actúa de forma diferente, inclusive artificial, cuando se le observa. Parece ser que los observadores ejercen muy poco efecto en las situaciones que observan. Se percibe que los individuos y los grupos se adaptan más bien rápidamente a la presencia de un observador y que actúan como normalmente lo harían. Esto no quiere decir que el observador no pueda ejercer un efecto. Quiere decir que si el observador es cuidadoso para no interferir y para evitar que las personas observadas sientan que se están haciendo juicios, entonces el observador, como estímulo influyente, es prácticamente anulado. Babbie (1995) afirma que no existe una protección completa para el efecto del observador. Sin embargo, el conocimiento y la sensibilidad ante este problema ofrecen una protección parcial.

Validez y confiabilidad

En la superficie, nada parece más natural cuando se observa el comportamiento, que creer que se está midiendo lo que se dice que se está midiendo. Sin embargo, cuando se da una carga interpretativa al observador, la validez puede verse afectada (así como la confiabilidad). A mayor carga de interpretación, mayor será el problema de la validez. No obstante, ello no significa que no deba darse una carga interpretativa al observador.

Un aspecto simple de la validez de las medidas de observación es su poder predictivo. ¿Predicen criterios relevantes de forma confiable? El problema, como siempre, reside en los criterios. Las medidas independientes de las mismas variables son infrecuentes. ¿Es posible afirmar que la medida de observación del comportamiento de un maestro es válida debido a que se correlaciona positivamente con las calificaciones de los superiores? Se podría tener una medida independiente sobre necesidades orientadas hacia uno mismo, ¿pero sería esta medida un criterio adecuado para la observación de dichas necesidades?

Una clave importante para el estudio de la validez de medidas de observación del comportamiento parece ser la validez de constructo. Si las variables que se miden a través de un método de observación están inmersas dentro de un marco teórico, entonces debe existir cierta relación. ¿Verdaderamente existen? Suponga que una investigación incluye la teoría de la autoeficacia de Bandura (1982), y que se ha construido un sistema de observación cuyo propósito es medir la competencia en el desempeño. En efecto, la teoría indica que la autoeficacia percibida, o la autopercepción de competencia, afecta la competencia del desempeño real de una persona: a mayor autoeficacia, mayor será la competen-

cia en el desempeño. Si se encuentra que la autopercepción de la competencia y las medidas de la competencia real observada al efectuar cierta tarea prescrita son positivas y altas, entonces se apoya la hipótesis derivada de la teoría. Sin embargo, esto también constituye evidencia de la validez de constructo del sistema de observación.

La confiabilidad de los sistemas de observación es un asunto simple, aunque por ningún motivo es fácil. Con frecuencia se define como el acuerdo entre observadores. A partir de este punto de vista, los registros de películas, videocintas y audiocintas ayudan a lograr una confiabilidad muy alta. No obstante, el acuerdo entre observadores tiene defectos potenciales. Por ejemplo, la magnitud de un índice de acuerdo en parte se debe al acuerdo por azar y, por lo tanto, requiere corrección. Quizás el camino más fácil a seguir sea utilizar diferentes métodos para evaluar la confiabilidad, tal como se haría con cualquier medida utilizada en investigación del comportamiento: acuerdo de los observadores, confiabilidad repetida y el método de análisis de varianza. La evaluación de la confiabilidad y el acuerdo entre observadores son problemas especialmente difíciles de la observación directa, ya que los estadísticos comunes dependen del supuesto de que las medidas son independientes —aunque con frecuencia no son independientes—. La mayor parte del trabajo realizado en esta área proviene o se basa en el trabajo de Cohen y Fleiss sobre el coeficiente kappa (Fleiss, 1986; Fleiss y Cohen, 1973). El uso de la teoría de la generalizabilidad es prometedora en la medición de la confiabilidad para datos nominales (Li y Lautenschlager, 1997). Algunos desarrollos provienen de las ciencias de la salud, donde la observación y el acuerdo desempeñan un papel importante en el análisis y en las decisiones (véase Dunn, 1989, 1992). Medley y Mitzel (1963) ofrecen una profunda, pero técnicamente compleja, exposición sobre la confiabilidad de valoraciones en un marco del análisis de varianza. Hollenbeck (1978) y Rowley (1976) discuten acerca de la confiabilidad de las observaciones cuando las medidas son nominales. Revisiones más recientes sobre la confiabilidad de las observaciones y la confiabilidad intercalificadora se encuentran en Dewey (1989), McDermott (1988), Perreault y Leigh (1989), Schouten (1986), Topf (1986), Zegers (1991) y Zwick (1988). El artículo de Perreault y Leigh trata el tema desde el punto de vista de la investigación de mercado. Topf revisa el uso de las medidas de confiabilidad nominales en investigación de enfermería, y McDermott analiza su aplicación en la psicología escolar. Chan (1987), Oud y Sattler (1984), y Powers (1985) han desarrollado programas computacionales para ayudar a los investigadores a calcular estadísticos del acuerdo entre observadores.

Entonces, es necesario definir con precisión y sin ambigüedades lo que se va a observar. Si se está midiendo la *curiosidad*, se debe decir al observador qué es un comportamiento curioso. Si se mide la *cooperatividad*, se debe explicar de alguna manera al observador de qué forma se distingue el comportamiento cooperativo de otros tipos de comportamiento. Esto quiere decir que se debe proporcionar al observador alguna forma de definición operacional de la variable que se está midiendo; la variable debe definirse en términos de comportamiento.

Categorías

La tarea fundamental del observador consiste en asignar categorías a los comportamientos. Recuerde, del trabajo previo sobre particiones, que las categorías deben ser exhaustivas y mutuamente excluyentes. Para satisfacer la condición de exhaustividad primero debe definirse U, el universo de comportamientos que se van a observar. En algunos sistemas de observación esto no es difícil de lograr. McGee y Snyder (1975), para comprobar la hipótesis de que la gente que sala sus alimentos antes de probarlos, percibe que el control del

comportamiento está dentro del individuo (control disposicional), más que en el ambiente (control situacional), simplemente observaron la conducta de salar la comida de los participantes en restaurantes. En otros sistemas de observación resulta más difícil. Muchos, o la mayoría de los sistemas de observación citados en la enorme antología de instrumentos de observación de Simon y Boyer (1970), *Mirrors for Behavior*, son complejos y difíciles de utilizar. ¡Este trabajo consiste de 14 volúmenes de instrumentos de observación del comportamiento! La mayoría de tales instrumentos, 67 de 79, son utilizados para observaciones educativas. Los lectores que pretendan utilizar observación de comportamiento en su investigación deben consultar dicha información, en especial el volumen 1, que contiene un análisis general en las páginas 1-24.

En concordancia con el énfasis de este libro —de que el propósito de la mayoría de la observación es medir variables— se cita un sistema de observación en el salón de clases del interesante y creativo trabajo de Kounin y sus colegas (Kounin y Doyle, 1975; Kounin y Gump, 1974). El sistema reportado es más complejo que el sistema de observación de salar y probar los alimentos, pero mucho menos complejo que muchos sistemas de observación en el salón de clases. La variable observada fue la de compromiso con la tarea, que se observó al videofilmar 596 lecciones y luego observarlas en video para obtener medidas de compromiso. Las medidas se clasificaron como alto compromiso con la tarea y bajo compromiso con la tarea. Los autores también midieron la continuidad en las lecciones creando categorías que reflejaban mayor o menor continuidad en las mismas. Cuando se observaba el comportamiento de niños, utilizaron las categorías para registrar los comportamientos pertinentes observados.

Unidades de comportamiento

Decidir qué unidades utilizar en la medición del comportamiento humano continúa siendo un problema sin resolver. Aquí con frecuencia se enfrenta un conflicto entre las demandas de la confiabilidad y la validez. Teóricamente es posible lograr un alto grado de confiabilidad utilizando unidades pequeñas y fáciles de observar y registrar. Puede intentarse definir el comportamiento de manera operacional al hacer una lista de un gran número de actos de comportamiento, pudiendo lograr así, por lo general, un alto grado de precisión y confiabilidad. Sin embargo, al hacer esto también puede reducirse tanto el comportamiento que ya no guarde demasiada semejanza con el comportamiento que se intentaba observar, con lo cual se pierde validez.

Por otra parte, pueden utilizarse amplias definiciones "naturales" y quizás lograr un alto grado de validez. Se podría instruir a los observadores para que observen la *cooperatividad* y definir el *comportamiento de cooperación* como "aceptar los métodos, sugerencias e ideas de otras personas; trabajar armónicamente con otros para lograr metas" o alguna definición más bien general. Si los observadores han tenido experiencia de grupo y comprenden los procesos grupales, entonces podría esperarse que pudieran evaluar de manera válida el comportamiento como cooperativo o no cooperativo, al utilizar dicha definición. Una definición tan general e incluso tan vaga como ésta le permite al observador capturar, si le es posible, toda la gama del comportamiento cooperativo. No obstante, su gran ambigüedad permite que se hagan diferentes interpretaciones, disminuyendo probablemente su confiabilidad.

Algunos investigadores que siguen un modelo fuertemente operacional insisten en que se realicen definiciones sumamente específicas de las variables observadas. Ellos enlistarán diversos comportamientos específicos que el observador debe anotar; ningún otro se observa ni se registra. Modelos extremos como éste pueden producir una confiabi-

lidad alta, pero también pueden perder parte del aspecto esencial de las variables observadas. Suponga que se hace una lista de 10 tipos específicos de comportamientos para *cooperatividad* y que el universo de posibles comportamientos consta de 40 o 50 tipos. En efecto, se perderán aspectos importantes de la cooperatividad. Aunque aquello que se mide puede medirse de forma confiable, quizá resulte bastante trivial o irrelevante, en parte, para la variable.

Cooperatividad

Éste es el problema molar-molecular de cualquier procedimiento de medición en las ciencias sociales. El *modelo molar* toma "todos" de comportamiento más grandes como unidades de observación. Unidades completas de interacción pueden especificarse como blancos de observación. El comportamiento verbal puede separarse en intercambios completos entre dos o más individuos, o en párrafos u oraciones completas. En contraste, el *modelo molecular* toma segmentos más pequeños de comportamiento como unidades de observación. Cada intercambio completo o parcial puede registrarse. Las unidades de comportamiento verbal quizá sean palabras o frases cortas. Los observadores molares empiezan con una variable general definida de forma amplia, como se dijo antes, y observan y registran una variedad de comportamientos bajo la única rúbrica. La interpretación que den al significado del comportamiento que observan depende de la experiencia y conocimiento que tengan. Por otra parte, los observadores moleculares tratan de sacar su propia experiencia, conocimiento e interpretación de la escena de observación; registran lo que ven y nada más.

Inferencia del observador

Los sistemas de observación difieren en otra importante dimensión: la *cantidad de inferencia* requerida del observador. Los sistemas moleculares requieren relativamente de poca inferencia. El observador simplemente anota si un individuo hace o dice algo. Por ejemplo, un sistema puede requerir que el observador anote cada unidad de interacción, que puede estar definida como cualquier intercambio verbal entre dos individuos. Si ocurre un intercambio, se anota; si no ocurre, no se anota. Otra categoría podría ser "golpea a otro niño" . Cada vez que un niño golpee a otro, ello se anota. No se hacen inferencias en este sistema —si, por supuesto, es acaso posible escapar a las inferencias (por ejemplo, "golpea")—. Se registra el comportamiento puro lo más posible.

Son escasos los sistemas de observadores con un nivel tan bajo de inferencia por parte del observador. La mayor parte de los sistemas requieren de cierto nivel de inferencia. Un investigador puede estar realizando investigación sobre el comportamiento del consejo de educación, y decide que un análisis con poca inferencia se ajusta al problema, y utiliza reactivos de observación como "sugiere un curso de acción", "interrumpe a otro miembro del consejo", "plantea una pregunta", "da una orden al superintendente", y otras similares. Puesto que dichos reactivos son ambiguos comparativamente, la confiabilidad de la observación necesita ser alta.

Los sistemas que requieren que el observador utilice altos niveles de inferencia son más comunes y probablemente más útiles en la mayor parte de la investigación. Los sistemas de observación de alta inferencia proveen al observador categorías denominadas, las cuales requieren de mayor o menor interpretación del comportamiento observado. Por ejemplo, suponga que se mide la *dominancia*, que se define como los intentos realizados por un individuo para mostrar superioridad intelectual (o de otro tipo) sobre otros indivi-

duos, con poco reconocimiento de las metas de grupo y de las contribuciones de otros. Esto, por supuesto, requerirá de un mayor nivel de inferencia del observador, y los observadores tendrán que entrenarse para que exista acuerdo sobre lo que constituyen comportamientos dominantes. Sin dicho entrenamiento y acuerdo —y probablemente sin experiencia en procesos grupales— la confiabilidad puede verse amenazada. Weick (1968) presenta una sofisticada exposición sobre la inferencia en la observación, y también analiza los sesgos en la observación y sugiere soluciones metodológicas para minimizar los efectos del sesgo. Señalamientos similares son pertinentes cuando se intentan medir muchas variables psicológicas y sociológicas: cooperación, competencia, agresividad, democracia, aptitud verbal, rendimiento y clase social, por ejemplo. Para revisar discusiones más recientes sobre observación e inferencia en estas áreas se recomienda leer Alexander, Newell, Robbins y Turner (1995); Borich y Klinzing (1984); Chavez (1984); Hartmann y Wood (1990); Jaffe (1997); Nurius y Gibson (1990), y Timberlake y Silva (1994). Los artículos de Borich y Klinzing y de Chavez se aplican a las observaciones en salón de clases. Hartmann y Wood tratan los sistemas de observación del comportamiento utilizados en modificación conductual. Nurius y Gibson tratan la observación e inferencia clínica encontrada en el trabajo social. En estrecha relación están los artículos de Jaffe y el de Alexander *et al.*, que tratan las observaciones clínicas. Timberlake y Silva tratan la observación e inferencia obtenida al observar la conducta de animales.

No es posible realizar generalizaciones uniformes sobre las virtudes relativas de sistemas con diferentes niveles de inferencia. Tal vez el mejor consejo para el neófito sea buscar un nivel medio de inferencia. Las categorías demasiado vagas, con muy poca especificación sobre qué se va a observar, ponen una carga excesiva sobre el observador. Es muy fácil que diferentes observadores den distintas interpretaciones al mismo comportamiento. Las categorías demasiado específicas, aunque reducen la ambigüedad y la incertidumbre, pueden tender a ser demasiado rígidas e inflexibles, e incluso triviales. Lo mejor es que el lector estudie varios sistemas exitosos, poniendo especial atención a las categorías de comportamiento y a las definiciones (instrucciones) ligadas a las categorías para guía del observador.

Generalización y aplicabilidad

Los sistemas de observación difieren considerablemente en su *generalización*, o en el grado de *aplicabilidad* a las situaciones de investigación distintas de aquellas para las que fueron diseñadas originalmente. Algunos sistemas son bastante generales: están diseñados para utilizarse con muchos problemas de investigación diferentes. El reconocido análisis de grupo de interacción de Bales (1951) es uno de dichos sistemas generales. Es un sistema de baja inferencia en el que todo el comportamiento verbal y no verbal, presuntamente en cualquier grupo, puede clasificarse dentro de una de 12 categorías: "muestra solidaridad", "está de acuerdo", "pide opinión", etcétera. Las 12 categorías están agrupadas dentro de tres grandes conjuntos: social emocional positivo, social emocional negativo y de tarea neutral.

Sin embargo, algunos sistemas fueron construidos para situaciones particulares de investigación, para medir variables particulares. El ejemplo anterior sobre salar los alimentos, es bastante específico y difícilmente se aplica a otras situaciones. El sistema de Kounin y Doyle (1975), que aunque fue construido específicamente para la investigación de Kounin, se aplica en muchas situaciones del salón de clases. De hecho, la mayor parte de los sistemas elaborados para problemas de investigación específicos se utilizan a menudo con ciertas modificaciones, en otros problemas de investigación.

Resulta necesario enfatizar que los "pequeños" sistemas de observación sirven para medir variables específicas. Suponga, por ejemplo, que la atención de los alumnos de escuela primaria sea una variable clave en una teoría sobre el rendimiento escolar. La atención (como rasgo o hábito), por sí mismo, ejerce poco efecto sobre el rendimiento: considere que la correlación es cero. Es una variable clave debido a que interactúa con cierto método de enseñanza y tiene un efecto pronunciado indirecto sobre el rendimiento. Asumiendo que esto es así, se debe medir la atención. Parece claro que se tendrá que observar el comportamiento del alumno, mientras se esté utilizando el método en cuestión y un método de "control". En tal caso, se necesita encontrar o diseñar un sistema de observación que se enfoque en la atención. Para evaluar la influencia del ambiente del salón de clases, por ejemplo, Keeves (1972) concluyó que era necesario medir la atención al observar a los estudiantes a quienes les pedía que pusieran atención a una tarea asignada por el maestro. Se asignaron puntuaciones que indicaban atención o la falta de ella. Este "pequeño" sistema de observación era confiable y aparentemente válido. Es probable que sistemas con objetivos específicos como éste incrementen su empleo en la investigación del comportamiento, especialmente en educación.

Muestreo del comportamiento

El muestreo, la última característica de las observaciones, estrictamente hablando no es una característica. Es una forma para obtener observaciones. Antes de usar un sistema de observación en investigación, debe decidirse cuándo y cómo se aplicará el sistema. Si se va a observar el comportamiento de los maestros en el salón de clase, ¿cómo se muestrearán los comportamientos? ¿Se observarán todos los comportamientos específicos en un periodo de clase? ¿O se harán muestreos de forma sistemática y aleatoria de comportamientos específicos? En otras palabras, debe diseñarse y utilizarse un plan de muestreo de algún tipo.

Existen dos aspectos del muestreo del comportamiento: el muestreo de eventos y el muestreo de tiempo. El *muestreo de eventos* es la selección de la observación de las ocurrencias integrales de comportamiento o eventos de cierta clase. Ejemplos de eventos integrales son los berrinches, las peleas y disputas, los juegos, los intercambios verbales sobre temas específicos, las interacciones entre alumnos y maestros en el salón de clases, etcétera. El investigador que estudia eventos debe saber cuándo ocurrirán dichos eventos y debe estar presente cuando sucedan, como con eventos del salón de clases; o esperar hasta que sucedan, como en las peleas.

El muestreo de eventos posee tres virtudes: 1) Los eventos son situaciones naturales de la vida y, por lo tanto, tienen una validez inherente que las muestras de tiempo por lo común no poseen. 2) Un evento integral posee una continuidad de comportamiento que los actos de comportamiento fragmentados de las muestras de tiempo no poseen. Si se observa una situación de solución de problemas desde el inicio hasta el final, entonces se está contemplando una unidad completa y natural de comportamiento individual y grupal. Al hacerlo, se logra una unidad completa y realista más grande del comportamiento individual y social. Como se estudió en un capítulo anterior, cuando se expusieron los experimentos de campo y los estudios de campo, las situaciones naturales impactan y se acercan a la realidad psicológica y social de una manera que los experimentadores normalmente no logran. 3) La tercera virtud del muestreo de eventos implica una característica importante de muchos eventos de comportamiento: en algunas ocasiones son inusuales y poco frecuentes. Por ejemplo, se puede estar interesado en las decisiones tomadas en reuniones administrativas y legislativas; o tal vez interesarse en el último paso de la solución de

problemas. Los métodos disciplinarios de los maestros constituyen una variable. Dichos eventos y muchos otros son relativamente poco frecuentes. Como tales, pueden perderse fácilmente por el muestreo de tiempo; por lo tanto, requieren de un muestreo de eventos. Sin embargo, si se toma el punto de vista más activo de observación promulgado por Weick (1968), es posible arreglar las situaciones para asegurarse de la ocurrencia más frecuente de eventos que suceden en pocas ocasiones.

El *muestreo de tiempo* es la selección de unidades de comportamiento para observación en diferentes momentos del tiempo. Pueden seleccionarse de formas sistemáticas o aleatorias para obtener muestras del comportamiento. Un buen ejemplo es el comportamiento del maestro. Suponga que se estudian las relaciones entre ciertas variables como el estado de alerta, la justicia y la iniciativa del maestro, por una parte; y la iniciativa y cooperación del alumno, por la otra. Se pueden seleccionar muestras aleatorias de maestros y después tomar muestras de tiempo de sus actos de comportamiento. Tales muestras de tiempo pueden ser sistemáticas: tres observaciones de 5 minutos en momentos específicos durante cada una de, por ejemplo, cinco horas de clase, siendo las horas de clase el primero, tercero y quinto periodos de un día, y el segundo y cuarto periodos del día siguiente. O pueden ser al azar: cinco periodos de 5 minutos de observación seleccionados aleatoriamente de un universo especificado de periodos de 5 minutos. De hecho existen muchas maneras de establecer y seleccionar muestras de tiempo. Como siempre, la forma en que se eligen dichas muestras, su duración y su número debe estar determinada por el problema de investigación. En un fascinante estudio sobre el liderazgo y el poder de la influencia grupal en niños pequeños, Merei (1949) señala que el muestreo de tiempo sólo mostraría líderes dando órdenes y al grupo obedeciendo; mientras que observaciones prolongadas mostrarían los mecanismos internos del hecho de dar órdenes y obedecer.

Las muestras de tiempo tienen la importante ventaja de incrementar la probabilidad de obtener muestras representativas de comportamiento. Sin embargo, ello es así sólo con los comportamientos que ocurren con mucha frecuencia. Los comportamientos poco frecuentes tienen una alta probabilidad de escapar de la red de muestreo, a menos que se seleccionen muestras enormes. El comportamiento creativo, compasivo y hostil, por ejemplo, quizá sea muy poco frecuente. Aun así, el muestreo de tiempo constituye una contribución positiva al estudio científico del comportamiento humano.

Como se explicó antes, las muestras de tiempo carecen de continuidad, de un contexto adecuado y, aun, de naturalidad. Lo anterior es cierto particularmente cuando se utilizan pequeñas unidades de tiempo y de comportamiento. Sin embargo, no hay razón para que el muestreo de eventos y el muestreo de tiempo no puedan combinarse algunas veces. Si se estudian recitaciones en el salón de clases, se selecciona una muestra aleatoria de los periodos de clase de un maestro en diferentes momentos, y se observan todas las recitaciones durante los periodos que se muestrearon, en su totalidad.

Algunas referencias muy buenas sobre el muestreo de eventos y el muestreo de tiempo se encuentran en Arrington (1943), Martin y Bateson (1993), Wright (1960) y Zeren y Makosky (1986). El artículo de Zeren y Makosky es sobresaliente, pues describe un ejercicio de salón de clases para enseñar a los estudiantes a realizar observaciones sistemáticas de comportamiento humano espontáneo. Además se describen tres técnicas de observación (muestreo de tiempo, muestreo de eventos y calificación de rasgos) y se compara su uso en comportamientos simulados presentados en televisión. La actividad en la clase incluyó una conferencia sobre métodos de observación, un ejercicio en donde se utilizaba uno de los tres métodos y una discusión en la clase. El artículo expone cómo enseñar el método científico para reunir datos de observación, la importancia de elaborar definiciones operacionales precisas para el acuerdo intercalificadores y el cálculo de coeficientes de confiabilidad.

Escalas de calificación

Hasta este punto, se ha hablado únicamente acerca de la observación de *comportamiento real*. Los observadores observan y escuchan directamente a los objetos en cuestión. Se sientan en el salón de clases y observan las interacciones maestro-alumno y alumno-alumno. O quizá vean y escuchen a un grupo de niños que resuelve un problema, frente a un espejo de doble vista.

Sin embargo, existe otra clase de observación del comportamiento que necesita mencionarse. Este tipo de observación se denominará *comportamiento recordado* o *comportamiento percibido*. Está clasificado convenientemente bajo el tema de las escalas de calificación. Para medir el comportamiento recordado o percibido, por lo común se les presenta a los observadores un sistema de observación en forma de escala de algún tipo, y se les pide que evalúen una o más características de un objeto, cuando el objeto no esté presente. Para hacerlo, ellos deben evaluar basándose en observaciones pasadas o en percepciones sobre cómo es el objeto observado y sobre cómo se comportará. Una forma conveniente para medir tanto el comportamiento real como el comportamiento percibido o recordado son las escalas de calificación.

Una *escala de calificación* es un instrumento de medición que requiere que un calificador u observador asigne al objeto calificado categorías o continuos que poseen valores numéricos asignados a ellos. Las escalas de calificación son quizás los instrumentos de medición más comunes, probablemente debido a que en apariencia son fáciles de construir y, lo más importante, son fáciles y rápidas de utilizar. Por desgracia, su aparente facilidad de construcción es engañosa, y la facilidad de uso conlleva un precio alto: la falta de validez debida a un número de fuentes de sesgo que entran en las medidas de calificación. No obstante, con conocimiento, habilidad y cuidado, las calificaciones resultan valiosas.

Para revisar un excelente estudio de las escalas de calificación, véase Guilford (1954), Nunnally (1978), Nunnally y Bernstein (1993) y Torgeson (1958). Si se desea revisar una presentación relativamente poco técnica sobre las escalas de calificación se recomienda leer a Selltiz, Jahoda, Deutsch y Cook (1961). Aunque las escalas de calificación ya se mencionaron con anterioridad en este libro, no se analizaron de manera sistemática. Al leer lo que sigue, el estudiante debe tener en mente que las escalas de calificación son en realidad escalas objetivas y, como tales, deberían haberse incluido en el capítulo 30. Su exposición se reservó para este capítulo a causa de que la exposición del capítulo 30 está enfocada principalmente en medidas donde responde el sujeto a quien se está midiendo. Las escalas de calificación, por otro lado, son medidas de individuos y sus reacciones, características y comportamientos, realizadas por observadores. Entonces, el contraste está en la forma en que el sujeto se observa a sí mismo y cómo lo perciben los demás. Las escalas de calificación también sirven para medir objetos, productos y estímulos psicológicos, tales como la escritura manual, los conceptos, los ensayos, los protocolos de entrevista y los materiales de pruebas proyectivas.

Tipos de escalas de calificación

Existen cuatro o cinco tipos de escalas de calificación, dos de los cuales se analizaron en el capítulo 30. Se trata de los listados y los instrumentos de elección forzada. Ahora se consideran sólo tres tipos y sus características: la escala de calificación de categorías, la escala de calificación numérica y la escala de calificación gráfica. Son bastante similares y difieren sólo en algunos detalles.

La *escala de calificación de categorías* presenta a los observadores o jueces varias categorías, de donde ellos eligen la que mejor caracteriza el comportamiento o características del objeto que se califica. Suponga que se califica el comportamiento de una maestra en el salón de clases. Una de las características que se califica es, por ejemplo, el estado de alerta. Un reactivo de categoría podría ser similar al que se mostró en el primer ejemplo. Una forma diferente utiliza descripciones condensadas; un reactivo de este tipo sería como el que se presenta en el segundo ejemplo

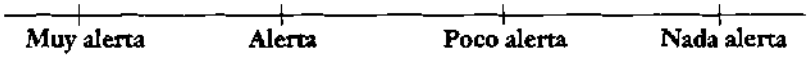
Ejemplos

¿Qué tan alerta es ella? (Marque una)
Muy alerta
Alerta
Poco alerta
Nada alerta

¿Emplea recursos? (Marque una)
Siempre emplea recursos; nunca le faltan ideas
Sus recursos son buenos
Algunas veces carece de ideas
No emplea recursos; rara vez tiene ideas

Las *escalas de calificación numérica* son, quizás, las más fáciles de construir y de utilizar. Además, también producen números que pueden usarse directamente en análisis estadísticos. Por otro lado, como los números representan intervalos iguales en la mente del observador, pueden alcanzar la medición intervalar (véase Guilford, 1954, p. 264). Cualquiera de las anteriores escalas de categorías puede convertirse fácil y rápidamente en escalas de calificación numérica, simplemente añadiendo números antes de cada una de las categorías. Los números 3, 2, 1, 0 o 4, 3, 2, 1 pueden agregarse al reactivo del *estado de alerta* mencionado anteriormente. Un método de calificación numérica conveniente consiste en el empleo del mismo sistema numérico, por ejemplo, 4, 3, 2, 1, 0 con cada reactivo. Éste es, por supuesto, el sistema utilizado en las escalas de actitud de puntuación sumada. Sin embargo, en las escalas de calificación, probablemente sea mejor dar tanto la descripción verbal como los valores numéricos.

En las *escalas de calificación gráfica* se combinan líneas o barras con frases descriptivas. El reactivo del estado de alerta, que se expuso antes, podría aparecer de la siguiente manera en forma gráfica:



Tales escalas incluyen muchas variedades: líneas verticales segmentadas, líneas continuas, líneas sin marcas, líneas divididas en intervalos iguales marcados (como la anterior) y otras. Probablemente se trata de las mejores formas de las escalas de calificación y las más utilizadas. Fijan un continuo en la mente del observador; sugieren intervalos iguales, y son claras y fáciles de comprender y de usar. Guilford (1954, p. 268) las sobrestima un poco cuando afirma: "son muchas las virtudes de las escalas de calificación gráfica, y sus fallas son pocas", pero su señalamiento se toma a bien.

Debilidades de las escalas de calificación

Las escalas de calificaciones tienen dos serias debilidades, una es extrínseca y la otra intrínseca. El defecto extrínseco consiste en que son aparentemente tan fáciles de construir

y de usar, que se utilizan de forma indiscriminada, a menudo sin conocimiento de sus defectos intrínsecos. No se hará una pausa para mencionar los errores que pueden escabullirse en la construcción y empleo inadecuados de las escalas de calificación. En lugar de eso, se alerta al lector en contra de su uso para cualquiera y todas las necesidades de medición. Primero debe plantearse la pregunta: ¿existe una mejor forma para medir mis variables? Si es así, es necesario utilizarla; si no, entonces se deben estudiar las características de las buenas escalas de calificación, trabajar con esmerado cuidado y poner los resultados de las calificaciones bajo prueba empírica y análisis estadístico adecuado.

El defecto intrínseco de las escalas de calificación es su tendencia al error constante o por sesgo, lo cual no es nuevo para el lector, por supuesto, pues dicho problema se abordó cuando se consideró la fijeza de respuesta. Sin embargo, con las calificaciones es especialmente amenazante para la validez. El error constante de calificación toma varias formas, de las cuales la más penetrante es el famoso *efecto de halo*. Se trata de la tendencia a valorar un objeto en la dirección constante de una impresión general del objeto. Casos diarios de halo son, por ejemplo, creer que un hombre es virtuoso porque nos agrada, y/o manifestar grandes elogios a los presidentes republicanos y condenar a los demócratas.

El halo se manifiesta con frecuencia en la medición, especialmente con las calificaciones. Los profesores evalúan más alta la calidad de las preguntas de una prueba de ensayo de lo que deberían, porque les simpatiza el examinado. O tal vez califiquen más alto (o más bajo) de lo que deberían la segunda, tercera y cuarta preguntas, debido a que la primera pregunta estuvo bien contestada (o mal contestada). La evaluación que hace el maestro del rendimiento de los niños, en la cual influye la docilidad o falta de docilidad de los niños, constituye otro caso de halo. Al evaluar a los individuos con escalas de calificación, existe la tendencia de que la calificación de una característica influya en las calificaciones de otras características. Es difícil evitar el halo. Parece ser particularmente fuerte en rasgos que no están claramente definidos, que no son fáciles de observar y que son moralmente importantes (véase Guilford, 1954, p. 279).

Dos fuentes importantes de error constante son el error por severidad y el error por flexibilidad. El *error por severidad* es la tendencia general de calificar demasiado bajo a todos los individuos, en todas las características. Es el del duro crítico: "nadie obtiene un 10 en mis cursos". El *error por flexibilidad* es la tendencia general opuesta de calificar demasiado alto. Éste es el caso del buen amigo que estima a todos, y la estimación se refleja en las calificaciones.

Una fuente exasperante de invalidez de las calificaciones es el *error de tendencia central*, que es la tendencia general a evitar todos los juicios extremos y a calificar justo en el centro de una escala de calificación. Se manifiesta particularmente cuando los calificadores no están familiarizados con los objetos que se están calificando.

Existen otros tipos de error de menor importancia que no se consideran aquí. Es más importante saber cómo enfrentar los tipos de error mencionados antes. Se trata de un tema complejo que no puede estudiarse aquí. Se refiere al lector a Guilford (1954, pp. 280-288, 383, 395-397), donde se discuten a detalle muchas estrategias para lidiar con el error. Los errores sistemáticos pueden tratarse en cierto grado por medio de las medias estadísticas. Guilford ha creado un método ingenioso que utiliza el análisis de varianza. La idea básica es que las varianzas debidas a los participantes, los jueces y las características se extraen de la varianza total de calificaciones; entonces se corrigen las calificaciones. Un método más fácil, cuando se califica una sola característica de los individuos, es el análisis de varianza de dos factores (grupos correlacionados). La confiabilidad también puede calcularse con facilidad. El uso del análisis de varianza para estimar la confiabilidad, como se aprendió antes, fue una contribución de Hoyt (1941). Ebel (1951) aplicó el análisis de varianza a la confiabilidad de las calificaciones.

Las escalas de calificación deben usarse en la investigación del comportamiento. Su uso no garantizado, expedito e inexperto ha sido justamente condenado. Pero esto no debe significar una condena general. Tienen virtudes que las hacen valiosas herramientas de investigación científica: requieren de menos tiempo que otros métodos, por lo común son interesantes y fácil de usar por los observadores, poseen un rango muy amplio de aplicación, y pueden utilizarse con un gran número de características. Hay que añadir el hecho de que es posible usarlas en conjunto con otros métodos, es decir, servirse de ellas como instrumentos de ayuda para las observaciones del comportamiento, y utilizarlas en conjunto con otros instrumentos objetivos, con entrevistas y aun con medidas proyectivas.

Ejemplos de sistemas de observación

Otros sistemas de observación del comportamiento (no mencionados antes) se resumen a continuación, para ayudar al estudiante a tener una idea de la variedad de sistemas posibles y de las maneras en que se construyen y utilizan dichos sistemas. Además, el lector comprenderá mejor cuándo es apropiada una observación del comportamiento.

Muestreo de tiempo del comportamiento de juego de niños con problemas auditivos

El comportamiento de juego se considera un componente importante del desarrollo normal del niño. No obstante, los niños con problemas auditivos tienen déficits de comunicación que interfieren con el desarrollo normal del juego. Se ha descubierto que los niños con problemas auditivos se involucran en juegos menos complejos y menos intercambio social, que los niños que no tienen estos problemas. En su estudio sobre el comportamiento de juego de niños con problemas auditivos, Esposito y Koorland (1989) utilizaron una técnica de muestreo de tiempo momentáneo para registrar el comportamiento de dos niños en diferentes lugares de juego (uno de 3.5 años y otro de 5 años). La meta era utilizar los datos para comparar el comportamiento de niños con problemas auditivos cuando se relacionan y cuando no se relacionan con niños sin problemas auditivos. Esto incluyó la observación y el registro del comportamiento de cada niño durante intervalos de 10 segundos en dos sesiones de 10 minutos por día, durante cuatro días a la semana, por dos semanas, durante el juego libre en interiores. Un ambiente estaba integrado y el otro no. Se codificó el comportamiento de juego libre de cada niño, de acuerdo con las categorías de juego definidas por Higginbotham, Baker y Neill (1980). Existen ocho categorías principales de juego que se clasifican, a su vez, en *juego social, juego cognitivo* y *sin juego*. Los investigadores encontraron diferencias en el comportamiento de juego entre los dos tipos de ambientes. Si la interacción de los compañeros durante el juego contribuye al desarrollo normal del niño, entonces sus resultados sugieren que los ambientes integrados son más adecuados para los niños con problemas auditivos.

Observación y evaluación de la enseñanza universitaria

En uno de los relativamente pocos —y mejores— estudios sobre los maestros y la enseñanza universitaria, Isaacson, McKeachie, Milholland y Lin (1964), después de mucho trabajo preliminar en los reactivos y sus dimensiones o factores, les pidieron a los estudiantes universitarios que calificaran y evaluaran a sus maestros, con base en el recuerdo de sus

observaciones e impresiones. Se ha publicado un número de estudios similares desde la aparición de dicho estudio; sin embargo, continúa siendo uno de los mejores. El sistema de observación presentado no se diseñó deliberadamente para medir variables, sino para ayudar a la evaluación del desempeño de los maestros. No obstante, sus dos dimensiones básicas pueden, por supuesto, utilizarse como variables de investigación. Un aspecto notable de los estudios que evalúan maestros universitarios es que los investigadores no parecen estar conscientes de que el propósito de dichos sistemas de observación debe ser la mejora de la instrucción (o que utilicen sus dimensiones como variables de investigación), y no propósitos de tipo administrativo (véase Kerlinger, 1971).

Isaacson *et al.* utilizaron una escala de calificación con 46 reactivos e instruyeron a los estudiantes a responder de acuerdo con la frecuencia de la ocurrencia de ciertos actos de comportamiento, y no de acuerdo con el hecho de si los comportamientos eran deseables o indeseables. Su interés básico se centraba en las dimensiones o variables subyacentes a los reactivos. Ellos encontraron seis de tales dimensiones (factores). La primera dimensión estaba relacionada con las habilidades generales de enseñanza.

Aunque los seis factores son importantes debido a que parecen mostrar diversos aspectos de la enseñanza (por ejemplo, la estructura, que es la organización que hace el instructor del curso y sus actividades; y el *rapport*, que es el aspecto más interactivo de la enseñanza y la amistad), el enfoque aquí será en el primer factor. A continuación se presentan tres de los reactivos:

> Él transmitía su material en una forma interesante. Él estimulaba la curiosidad intelectual de sus estudiantes. Él daba explicaciones claras y sus explicaciones iban al grano (p. 347).

Sin embargo, el reactivo más efectivo era incluso más general:

¿Cómo calificarías a tu instructor respecto a su habilidad general (completa) de enseñanza?

a) Un instructor sobresaliente y estimulante
b) Un instructor muy bueno
c) Un buen instructor
d) Un instructor adecuado, pero no estimulante
e) Un instructor pobre e inadecuado

Mientras que es posible cuestionar el hecho de denominar a este estudio y a otros parecidos como estudios de observación, sí existe cierta observación, aunque es bastante diferente al tener que recordarse y ser indirecta, global y altamente inferencial y, finalmente, mucho menos sistemática que la observación real. Se pide a los estudiantes que recuerden y califiquen comportamientos a los cuales pueda no haber puesto especial atención. No obstante, el estudio de Isaacson *et al.*, aunado a otros estudios, demostró que esta forma de observación puede utilizarse en la confiabilidad de la evaluación del instructor y del curso.

Evaluación de la observación del comportamiento

No cabe duda de que la observación objetiva del comportamiento humano ha avanzado más allá de la etapa rudimentaria. Los avances, al igual que otros avances metodológicos y de medición logrados en los pasados 10 a 20 años, han resultado sorprendentes. El incremento del dominio y sofisticación psicométricos y estadísticos se manifiestan en la observación y evaluación del comportamiento real y recordado. La investigación científica

social puede beneficiarse —y efectivamente lo hará— con estos avances. Muchos problemas de investigación educativa, por ejemplo, demandan enérgicamente observaciones del comportamiento: los niños en salones de clases que interactúan entre sí y con sus maestros, administradores y maestros que discuten problemas escolares en juntas de personal, consejos de educación que trabajan sobre decisiones políticas, etcétera. Tanto la investigación básica como la aplicada, en especial la investigación que involucra procesos y decisiones grupales, se benefician de la observación directa. Además puede utilizarse en estudios de campo, experimentos de campo y experimentos de laboratorio. He aquí un modelo metodológico que es esencialmente igual en las situaciones de campo y de laboratorio.

La dificultad del uso de sistemas de escala completa sin duda ha desmotivado el uso de la observación en la investigación del comportamiento. No obstante, las observaciones deben usarse cuando las variables de estudios de investigación sean de naturaleza interactiva e interpersonal, y cuando se desee estudiar las relaciones entre el comportamiento real, como las técnicas de manejo de clase o las interacciones grupales, y otros comportamientos o variables atributivas. Aunque es importante preguntar acerca del comportamiento, no hay un sustituto para observar, de forma tan directa como sea posible, lo que en realidad hace la gente cuando se enfrenta con diferentes circunstancias y diferentes personas. Además, quizá no sea necesario usar los sistemas de observación más grandes en la mayor parte de la investigación del comportamiento. Como se señaló anteriormente, es posible diseñar sistemas pequeños para propósitos específicos de investigación. El sistema limitado de Keeves (1972) mostró ser sumamente apropiado para sus propósitos. En cualquier caso, la investigación científica del comportamiento requiere de observaciones directas e indirectas del comportamiento, y los medios técnicos para la realización de dichas observaciones se están volviendo cada vez más adecuados y disponibles. En el siglo XXI se debe lograr una comprensión y mejoría considerables de los métodos de observación, así como el incremento de su uso significativo.

Sociometría

Constantemente se evalúa a la gente con quien uno trabaja, con quien va a la escuela y con quien vive en casa. Se juzga su adecuación para el trabajo, para el juego y para vivir con nosotros. Además los juicios se basan en nuestras observaciones sobre su comportamiento en distintas situaciones. Se dice que juzgamos con base en nuestra "experiencia". La forma de medición que ahora se considera sociometría está basada en muchas de esas observaciones informales. Nuevamente, el método se basa en observaciones recordadas y en los juicios inevitables que se hacen de la gente, después de observarla.

Sociometría y elección sociométrica

Sociometría es un término amplio que indica un número de métodos de reunión y análisis de datos sobre los patrones de elección, comunicación e interacción de los individuos de grupos. Se podría decir que la sociometría es el estudio y medición de la elección social. También se ha considerado como un medio para estudiar las atracciones y repulsiones de los miembros de un grupo.

Se le pide a una persona que elija a una o más personas, de acuerdo con uno o más criterios establecidos por el investigador: ¿con quién le gustaría trabajar? ¿Con quién le gustaría jugar? Entonces, la persona elige una, dos, tres o más opciones de entre los miembros de su propio grupo (generalmente) o de otros grupos. ¿Qué podría ser más simple y

natural? El método funciona bien tanto para los miembros de jardín de niños como para científicos atómicos.

La elección sociométrica debe comprenderse ampliamente: no significa tan sólo "elección personas", también puede significar "elección de líneas de comunicación", "elección de líneas de influencia" o "elección de grupos minoritarios". Las elecciones dependen de las instrucciones y preguntas dadas a los individuos. A continuación se presenta una muestra de una lista de preguntas e instrucciones sociométricas:

Ejemplo

¿Con quién le gustaría trabajar (jugar, sentarse, etcétera)?
¿Quiénes son los miembros de este grupo (grupo de edad, clase, club, por ejemplo) que a usted le agradan más (menos)?
¿Quiénes son los tres mejores (peores) alumnos de su clase?
¿A quién elegiría para representarlo en un comité para mejorar el bienestar de la facultad?
¿Quiénes son los cuatro individuos que tienen el mayor prestigio en su organización (clase, compañía, equipo)?
¿Cuáles son los dos grupos de gente más aceptables (menos aceptables) para usted como vecinos (amigos, socios de negocios, socios profesionales)?

En efecto, existen muchas posibilidades. Algunas de ellas se analizan en Lindzey y Byrne (1968). Además, tales posibilidades llegan a multiplicarse simplemente al preguntar: ¿quién piensa usted que lo elegiría para...? Y ¿quién piensa usted que el grupo elegiría para...? También se pide a los participantes que ordenen por rangos a otros utilizando criterios sociométricos, siempre y cuando no sean demasiados a ordenar; o se pueden utilizar escalas de calificación.

Si se solicita a los miembros de un grupo u organización que se califiquen entre sí utilizando uno o más criterios, las instrucciones sociométricas quedarían de manera similar a la siguiente:

Ejemplo

Aquí hay una lista de los miembros de su grupo. Califique a cada uno de acuerdo con el hecho de si a usted le gustaría trabajar con él en un comité encargado de establecer un conjunto de estatutos. Utilice los números 4, 3, 2, 1 y 0, donde 4 signifique que le gustaría mucho trabajar con él, 0 que no desearía trabajar con él en lo absoluto, y los otros números representan grados intermedios de cuánto le agradaría trabajar con él.

Evidentemente es posible utilizar otros métodos de medición. La principal diferencia radica en que la sociometría siempre implica ideas como la de elección, interacción, y comunicación sociales, y las influencias que están detrás de ellas.

Métodos de análisis sociométrico

Existen tres formas básicas de análisis sociométrico: matrices sociométricas, sociogramas o gráficas dirigidas e índices sociométricos. De todos los métodos de análisis sociométrico, las *matrices sociométricas,* que vamos a definir, quizás contengan las posibilidades e implicaciones más importantes para el investigador del comportamiento. Los *sociogramas* son diagramas o tablas de las elecciones realizadas en los grupos. Los sociogramas o gráficas dirigidas se tratarán muy poco, ya que se utilizan más para propósitos prácticos que de investigación, y su análisis matemático es difícil y requiere de mayor espacio del que se le

puede brindar aquí. Al lector que requiera mayor detalle y explicación se le recomienda consultar a Fienberg y Wasserman (1981) y Ove (1981). Los *índices sociométricos* son números sencillos calculados a partir de dos o más números producidos por datos sociométricos. Indican características sociométricas de individuos y grupos.

Matrices sociométricas

Se aprendió antes que una *matriz* es una tabla o arreglo rectangular de números o de otros símbolos. Para aquellos que no estén familiarizados con las matrices, se recomienda la lectura de Lindzey y Byrne (1968, pp. 470-473), donde encontrará un buen repaso sobre el análisis de matrices. En Kemeny, Snell y Thompson (1966, pp. 217-250, 384-406) se presentan explicaciones de operaciones elementales de matrices y matrices sociométricas. Una buena exposición elemental de matrices puede encontrarse en Davis (1973). Una revisión antigua, pero aun valiosa, de métodos matemáticos y estadísticos para el análisis de la estructura y comunicación de grupo es la que presentan Glanzer y Glaser (1959).

La sociometría casi siempre está relacionada con matrices cuadradas o de $n \times n$, donde n es igual al número de personas en un grupo. Los renglones de la matriz se denominan i, las columnas se denominan j. Por supuesto, i y j pueden representar cualquier número y cualquier persona en el grupo. Si se escribe a_{ij}, significa un dato en el renglón i y en la columna j de la matriz, o, de forma más simple, cualquier dato en la matriz. Es conveniente escribir *matrices sociométricas*, que son matrices de números que expresan todas las elecciones de los miembros de cualquier grupo.

Suponga que un grupo de cinco miembros respondió a la pregunta sociométrica "¿con quién le gustaría trabajar en tal o cual proyecto durante los próximos dos meses? Elija dos individuos". Las respuestas a la pregunta sociométrica son, evidentemente, *elecciones*. Si un miembro del grupo elige a otro miembro, la elección se representa por un 1. Si un miembro del grupo no elige a otro, la falta de elección está representada por un 0. (Si se hubiera incluido el rechazo, se podría utilizar –1.) La matriz sociométrica de elecciones, C, de esta situación grupal hipotética se presenta en la tabla 31.1.

Es posible analizar la matriz de varias formas. Pero primero es necesario asegurarse de saber cómo leer la matriz. Probablemente resulte más fácil leerla de izquierda a derecha, de i a j. El miembro i elige (o no elige) al miembro j. Por ejemplo, a elige b y e; c elige d y e. Algunas veces es conveniente expresarse en voz pasiva, "b fue elegido por a, d y e" o "c no fue elegido por ninguno".

TABLA 31.1 *Matriz sociométrica de elección: grupo de cinco miembros, pregunta de dos elecciones*[a]

		j				
		a	***b***	***c***	***d***	***e***
	a	0	1	0	0	1
	b	1	0	0	0	1
i	***c***	0	0	0	1	1
	d	0	1	0	0	1
	e	1	1	0	0	0
	Σ	2	3	0	1	4

[a] El individuo i elige al individuo j. Es decir, la tabla se lee por renglones: b elige a y e. También puede leerse por columnas: b es elegido por a, d y e. Las sumas de la sección inferior indican el número de elecciones que recibe cada individuo.

El análisis de una matriz por lo común se inicia observando quién elige a quién. Con una matriz simple esto es fácil. Existen tres tipos de elecciones: simple o de un factor, mutua o de dos factores y sin elección. Se analizarán primero las elecciones simples. (Esto se explicó en el párrafo anterior.) Una elección *simple* de un factor es cuando *i* elige *j*, pero *j* no elige *i*. En la tabla 31.1, *c* elige *d*, pero *d* no eligió *c*. Se escribe: $i \rightarrow j$, o $c \rightarrow d$. Una elección *mutua* es donde *i* elige *j* y *j* también elige *i*. En la tabla, *a* elige *b* y *b* elige *a*. Se escribe $i \leftrightarrow j$ o $a \leftrightarrow b$. Se podrían contar las elecciones mutuas en la tabla 31.1: $a \leftrightarrow b$, $a \leftrightarrow e$, $b \leftrightarrow e$.

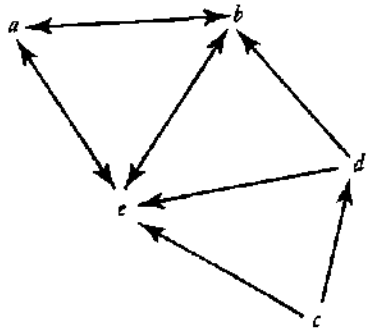
El grado en que cualquier miembro es elegido se percibe fácilmente al sumar las columnas de la matriz. En efecto, *e* es "popular": esta persona fue elegida por los demás miembros del grupo; *a* y *b* fueron elegidos dos y tres veces, respectivamente. De hecho *c* no es popular en lo absoluto: nadie eligió a esta persona; *d* tampoco es popular, ya que fue elegida sólo una vez. Si se les permite a los individuos un número ilimitado de elecciones, es decir, si se les instruye para elegir cualquier número de individuos, entonces las sumas de renglones cobran significado. Se les puede decir a los participantes que elijan una, dos, tres o más personas. Parece ser que tres es un número común de elecciones. El número permitido debe determinarse con base en los propósitos de investigación. Estas sumas se denominarían, por ejemplo, índices de *gregarismo*.

Existen otros métodos de análisis de matriz que son potencialmente útiles para los investigadores. Por ejemplo, por medio de operaciones de matrices relativamente simples se pueden determinar puntos y cadenas de influencia en grupos pequeños y grandes. Sin embargo, dichos aspectos van más allá del alcance de esta obra.

Sociogramas o gráficas dirigidas

Los análisis más simples son similares a aquellos que se acaban de exponer. Pero con una matriz más grande que la de la tabla 31.1 es casi imposible asimilar las complejidades de las relaciones de elección. Para esto, los *sociogramas* son útiles, siempre y cuando el grupo no sea demasiado grande. Ahora se cambia el nombre de "sociograma" por el de "gráfica dirigida", que es un término matemático más general que se aplica a cualquier situación donde *i* y *j* estén en alguna relación *R*. En lugar de decir "*i* elige *j*", es muy posible decir "*i* influye en *j*", "*i* se comunica con *j*", "*i* es amigo de *j*", o "*i* domina *j*". En lenguaje simbólico, se escribe: *iRj*. De manera específica, de los ejemplos expuestos arriba, se escribe: *iCj* (*i* elige *j*), *iIj* (*i* influye en *j*), *iCj* (*i* se comunica con *j*), *iFj* (*i* es amigo de *j*), *iDj* (*i* domina *j*). Cualquiera de estas interpretaciones puede describirse por una matriz como la de la tabla 31.1 y por medio de una gráfica dirigida. En la figura 31.1 se presenta una gráfica dirigida.

FIGURA 31.1



Con una simple ojeada se percibe que *e* es el centro de elección. A esta persona se le llamaría *líder*; o se diría que se trata de una persona agradable o competente. Aún más importante, note que *a*, *b* y *c* se eligieron entre sí. Conforman un eslabón. Un *eslabón* se define como tres o más individuos que se eligen entre sí. Festinger, Schachter y Back (1950) presentan un valioso método para identificar eslabones dentro de grupos. Al buscar más flechas dobles no se encuentra ninguna otra. Ahora se buscan individuos que no tengan flechas apuntando hacia ellos: *c* es uno de dichos individuos. Se afirma que *c* no es elegido, o que es rechazado. Si se desea más información sobre los eslabones y su identificación, consulte a Glanzer y Glaser (1959, pp. 326-327), quienes esbozan de forma sucinta métodos de la multiplicación de matrices binarias (1, 0), cuya aplicación ofrece ideas útiles respecto a la estructura grupal.

Observe que las gráficas dirigidas y las matrices dicen lo mismo. El número de elecciones que *a* recibe se observa añadiendo los números 1 en la columna *a* de la matriz. Se obtiene la misma información añadiendo el número de puntas de flecha que señalan hacia *a* en la gráfica. Para grupos de tamaño pequeño y mediano, y para propósitos descriptivos, las gráficas constituyen medios excelentes para sintetizar relaciones grupales. No son tan adecuadas para grupos más grandes (mayores de 20 miembros) ni para propósitos más analíticos, ya que se tornan difíciles de construir y de interpretar. Además, diferentes individuos pueden obtener gráficas diferentes utilizando los mismos datos. Las matrices son generales y, si se manejan apropiadamente, no son tan difíciles de interpretar. Con los mismos datos, diferentes individuos deben escribir exactamente las mismas matrices.

Índices sociométricos

En sociometría son posibles muchos índices. A continuación se muestran tres de ellos. El lector encontrará otros en la literatura. La presente exposición se basa, en su mayoría, en Proctor y Loomis (1951).

Un índice simple pero útil es:

$$EE_j = \frac{\sum c_j}{n-1} \tag{31.1}$$

donde EE_j = el estado de elección de la persona j; $\sum c_j$ = la suma de elecciones en la columna j; y n = el número de individuos en el grupo (se utiliza $n - 1$ porque no se puede contar al propio individuo). Para el sujeto *E* de la tabla 31.1, $CS_E = 4/4 = 1.00$ y para el sujeto $EE_a = 2/4 = .50$. *EE* revela qué tanto o qué tan poco es elegido un individuo; en pocas palabras, se trata del *estatus de elección*. Desde luego, es posible tener un índice de rechazo de elección. Simplemente se pone el número de los 0 en cualquier columna en el numerador de la ecuación 31.1.

Las medidas sociométricas grupales quizás sean más interesantes. Una medida de la cohesión de un grupo es:

$$Co = \frac{\sum (i \leftrightarrow j)}{\left[\frac{n(n-1)}{2}\right]} \tag{31.2}$$

La cohesión de grupo está representada por *Co* y $\sum(i \leftrightarrow j)$ igual a la suma de elecciones mutuas (o pares mutuos). Este útil índice es la proporción de las elecciones mutuas del número total de pares posibles. En un grupo de cinco miembros, el número total de pares posibles son cinco cosas, tomando dos a la vez:

$$\left\lfloor \frac{5}{2} \right\rfloor = \frac{5(5-1)}{2}$$

Si en una situación de elección ilimitada hubiera dos elecciones mutuas, entonces $C_o = 2/10 = .20$, un grado más bien bajo de cohesión. En el caso de elección limitada, la fórmula es:

$$Co = \frac{\Sigma(i \leftrightarrow j)}{\left[\frac{dn}{2}\right]} \tag{31.3}$$

donde d es igual al número de elecciones permitidas a cada individuo. Para C de la tabla 31.1 $Co = 3/(2 \times 5/2) = 3/5 = .60$, un grado sustancial de cohesión.

Usos de la sociometría en investigación

Debido a que los datos de la sociometría parecen ser tan diferentes a otros tipos de datos, los estudiantes tal vez encuentren difícil considerar la medición sociométrica como medición. No cabe duda de que los datos sociométricos son diferentes; pero son el resultado de observación y *son medidas*. Puesto que son medidas, también tienen los mismos problemas básicos de la medición, como la validez y la confiabilidad. Lindzey y Byrne (1968) analizan dichos aspectos de medición. Las mediciones sociométricas son útiles, por ejemplo, para clasificar individuos y grupos. En el estudio clásico del Bennington College, Newcomb (1943) midió el prestigio individual al pedir a los estudiantes que nombraran cinco estudiantes a los que ellos elegirían como los más valiosos para representar al Bennington College en una importante reunión de estudiantes, de todo tipo de universidades estadounidenses. El autor después agrupó a los estudiantes de acuerdo con la frecuencia de elección y relacionó esta medida de *prestigio sociométrico* con el conservadurismo político y económico. Al leer los ejemplos de tal sección, el estudiante debe comprender con claridad que la sociometría es un método de observación y recolección de datos que, como cualquier otro método de observación, obtiene medidas de variables.

Prejuicio en las escuelas

Rooney-Rebeck y Jason (1986) investigaron los efectos de la tutoría de compañeros de grupo cooperativo sobre las relaciones interétnicas de niños estadounidenses negros, blancos e hispanos de primero y tercer grados. Realizaron observaciones directas de las interacciones sociales en el patio de juegos, antes y después de un programa de intervención de ocho semanas. Los índices sociométricos se calcularon para medir las asociaciones interétnicas. Rooney-Rebeck y Jason encontraron un incremento en las interacciones interétnicas y en las elecciones sociométricas en los niños de primer grado, quienes también mostraron un incremento en sus calificaciones en aritmética y lectura. Sin embargo, no se encontraron cambios significativos entre los niños de tercer grado respecto a sus asociaciones interétnicas ni a su desempeño académico. A partir de tales resultados, parece ser que una estructura cooperativa de tutoría de compañeros en el salón de clases tiene beneficios para niños de primer grado, pero no necesariamente para niños de tercer grado. Los investigadores sugieren que ello puede deberse a una experiencia limitada en competencia académica y en prejuicios étnicos manifiestos, en los niños de primer grado.

Sociometría y estereotipos

Gross, Green, Storck y Vanyur (1980) utilizaron una combinación de calificaciones sociométricas y de estereotipos para estudiar las actitudes de las personas. Participantes de uno y otro sexo observaron una filmación, ya sea de un estímulo de una mujer homosexual o de un estímulo de un hombre homosexual. Los participantes se dividieron en tres grupos. A uno de los grupos de participantes se le informó, antes de ver la filmación, que la persona era homosexual. Al segundo grupo se le informó lo mismo, pero después de haber visto la filmación, y al tercer grupo no se le informó nada. Los resultados revelaron que las calificaciones de los rasgos eran más estereotípicas y las calificaciones sociométricas menos favorables para la persona del estímulo en ambas condiciones de revelación: inmediata o retardada. Aquellas personas identificadas como homosexuales fueron juzgadas de manera más estereotipada por los participantes del mismo sexo. Los hombres, por lo general, calificaron con mayor dureza a la persona en la situación retardada que en las condiciones de revelación inmediata.

Sociometría y estatus social

Se han realizado diversos estudios utilizando la sociometría para medir el estatus social. A continuación se presenta uno de ellos, el cual involucra el estatus social entre niños en edad escolar.

Inderbitzen, Walters y Bukowski (1997) estudiaron la relación entre grupos de estatus sociométrico y la ansiedad social en las relaciones entre compañeros adolescentes. Los participantes consistieron en 973 estudiantes del sexto grado de primaria al tercer grado de secundaria. El número de niños y niñas era casi igual. Los participantes completaron la *escala de ansiedad social para adolescentes* (Social Anxiety Scale for Adolescents) y una tarea de nominación sociométrica, la cual incluía descripciones comportamentales como el que más me gusta, el que menos me gusta, el que comienza las peleas la mayoría de las veces, el del mejor sentido del humor, el líder de la clase, el más fácil de influenciar y el más cooperador. Las nominaciones sociométricas se utilizaron después para clasificar a los estudiantes en grupos de estatus sociométrico estándar, tales como popular, promedio, rechazado, no tomado en cuenta y polémico; así como en subgrupos de rechazo: agresivo rechazado o sumiso rechazado. Los resultados indicaron que los estudiantes clasificados como rechazados y no tomados en cuenta reportaron mayor ansiedad social que aquellos clasificados como promedio, populares o polémicos. Además los estudiantes sumisos rechazados reportaron significativamente mayor ansiedad social que los estudiantes agresivos rechazados o promedio. A través del uso de la sociometría es posible descubrir la existencia de problemas entre compañeros adolescentes.

Raza, creencia y elección sociométrica

Graham y Cohen (1997) estudiaron la relación entre raza y sexo en las relaciones entre niños compañeros. Las relaciones se midieron a través de calificaciones sociométricas y amistades observadas. Dicho estudio observó a cada estudiante en una sola escuela primaria, incluyendo del 1o. al 6o. grados. La escuela fue elegida a causa de que tenía casi igual número de niños estadounidenses negros y blancos en cada clase. Además, los niños fueron divididos en dos grupos de edad: del 1o. al 3er. grados y del 4o. al 6o. grados. Independientemente de la edad, raza o sexo y de las medidas de relación, los niños favorecieron a los compañeros del mismo sexo más que a los compañeros de la misma raza, lo cual indica que los niños prefirieron las interacciones sociales con otros niños, sin importar la raza, y que las niñas prefirieron interactuar con otras niñas, sin importar la raza. Aunque los niños negros mayores tuvieron más amigos mutuos de la misma raza que de otra raza, los niños negros se mostraron más aceptantes de los niños blancos, que a la inversa. A

pesar de que hubo algunas preferencias por la misma raza, las evaluaciones entre razas fueron, por lo general, bastante positivas en ambas medidas de las relaciones entre compañeros.

La sociometría es un método simple, económico y natural de observación y de recolección de datos. Siempre que se involucren actos humanos, tales como la elección, la influencia, la dominación y la comunicación —especialmente en situaciones de grupo— pueden utilizarse métodos sociométricos, ya que poseen considerable flexibilidad. Si se les define de forma general, se adaptan a una amplia variedad de investigaciones tanto en el laboratorio como en el campo. Sus posibilidades de cuantificación y de análisis, aunque no siempre percibidas en la literatura, son recompensantes. La habilidad para utilizar la simple asignación de los 1 y los 0 es particularmente afortunada, debido a que pueden aplicarse métodos matemáticos poderosos a los datos, con resultados interpretables y significativos únicos. Los métodos de matrices constituyen el ejemplo sobresaliente, pues con ellos se descubren eslabones en grupos, canales de comunicación e influencia, patrones de cohesión, niveles de conexión, jerarquización, etcétera.

Como se mencionó antes, los métodos sociométricos, al igual que otros métodos, no carecen de defectos. Longshore (1982), por ejemplo, recomienda el empleo de otros métodos no intrusivos en lugar de la sociometría, cuando se estudian problemas delicados como la disgregación. Señala que los científicos sociales no han logrado hallazgos consistentes sobre la disgregación, a causa de que los investigadores se han preocupado más por los resultados de la disgregación, que por el amplio rango de condiciones bajo las cuales ocurre dicho fenómeno. Longshore considera que la evaluación de resultados a corto plazo debe evaluarse a través de medidas como la observación no intrusiva de los grupos de juego o las clases, en lugar de la sociometría.

RESUMEN DEL CAPÍTULO

1. Existe mucha controversia y debate respecto a la observación y a los métodos de observación.
2. Dos modelos básicos de observación son los siguientes:
   - Se observan los comportamientos abiertos de la gente (por ejemplo, lo que hacen y lo que dicen)
   - Se pregunta a la gente sobre sus propias acciones y sobre el comportamiento de otros
3. El observador puede representar un problema importante en el estudio. El observador tal vez haga inferencias incorrectas acerca del comportamiento observado.
4. El observador llega a afectar a los objetos de observación al formar parte del sistema observacional.
5. Las mediciones observadas están sujetas a requisitos de validez y confiabilidad. Al requerir que el observador interprete la observación, puede reducirse la validez.
6. La confiabilidad para las observaciones toma la forma de acuerdo entre jueces u observadores. El comportamiento que se observará a través de la observación directa debe establecerse claramente, con buenas definiciones operacionales.
7. Una tarea fundamental del observador consiste en categorizar los datos observados. Se crean categorías y, conforme se observan ciertos comportamientos, se hace una marca o nota en esa categoría.
8. Las unidades de comportamiento a veces son vagas o muy generales. Algunos investigadores utilizan definiciones operacionales muy estrictas.

9. Toda observación requiere de cierto nivel de interpretación por parte del observador.
10. Los diferentes sistemas de observación varían en su grado de generalización. Algunos son muy generales, mientras que otros son bastante específicos.
11. Los comportamientos pueden muestrearse con las técnicas del muestreo de eventos o del muestreo de tiempo. Cada una posee ventajas y desventajas.
12. Un tipo de observación implica presentar a los observadores con un sistema de observación en forma de escala de calificación. Se les pide que evalúen un objeto en términos de una o más características.
13. Existen cinco tipos de escalas de calificación:
    - listas de chequeo
    - instrumentos de elección forzada
    - escala de calificación de categorías
    - escala de calificación numérica
    - escala de calificación gráfica
14. Las escalas de calificación presentan dos serias debilidades:
    a) Puesto que son fáciles de construir y de utilizar, pueden crearse sin el conocimiento de sus defectos intrínsecos.
    b) Constantemente tienden al error o al sesgo.
15. La sociometría es un término amplio que indica el número de métodos de recolección y análisis de datos sobre patrones de elección, comunicación e interacción de individuos en grupos.
16. Existen tres formas básicas de análisis sociométrico: matrices sociométricas, sociogramas o gráficas dirigidas, e índices sociométricos.

SUGERENCIAS DE ESTUDIO

1. El lector debe estudiar en detalle uno o dos sistemas de observación del comportamiento. Para los alumnos de educación, el sistema de Medley y Mitzel (1963) producirá altos beneficios. Otros lectores desearán estudiar uno o dos sistemas distintos a éste, ya que el artículo tiene autoridad y claridad, y contiene muchos ejemplos. Las dos mejores referencias generales son las de Heyns y Lippitt (1954), y la de Weick (1968) en la primera y segunda ediciones del *Handbook of Social Psychology*. Una antología de 79 sistemas de observación se ha publicado por Simon y Boyer (1970) en cooperación con Research for Better Schools, Inc., un laboratorio regional de educación. El investigador que tenga la intención de realizar observaciones debe consultar esta amplia colección de sistemas. El estudiante de educación encontrará excelentes resúmenes y presentaciones sobre sistemas de observación educativa en Dunkin y Biddle (1974). Los siguientes artículos son valiosos. Boice señala la falta de entrenamiento para realizar observaciones del comportamiento y ofrece sugerencias para alcanzar dicho entrenamiento. Herbert y Attridge proporcionan criterios para los sistemas de observación. También señalan que el conocimiento de dichos sistemas es limitado.

Boice, R. (1983). Observational Skills. *Psychological Bulletin, 93*, 3-29.
Herbert, J. y Attridge, C. (1975). A guide for developers and users of observation systems and manuals. *American Educational Research Journal, 12*, 1-20.

2. Un investigador que estudia los patrones de influencia de los consejos de educación obtuvo la siguiente matriz, de un consejo de educación. (Note que es similar a una situación de elección ilimitada, ya que cada individuo puede influir en todos o en ninguno de los miembros del grupo.) La matriz se lee: *i* influye en *j*.

		j				
		a	*b*	*c*	*d*	*e*
	a	0	0	1	1	0
	b	0	0	0	0	1
i	*c*	1	0	0	1	0
	d	1	0	1	0	0
	e	0	1	0	0	0

   a) ¿Qué conclusiones se obtienen del estudio de esta matriz? ¿ El consejo está dividido? ¿Existe la posibilidad de que haya un conflicto?
   b) Dibuje una gráfica de la situación de influencia e interprétela.
   c) ¿Existe algún eslabón en el consejo? (Defina eslabón como se hizo en el texto.) Si es así, ¿quiénes son sus miembros?
   d) ¿Cuáles miembros poseen el menor número de canales de influencia? ¿Son ellos, entonces, mucho menos influyentes que los otros miembros, siempre y cuando lo demás se mantenga igual?
   [Respuestas: *c*) Sí: a, c, d; *d*) b y e.]

3. En el ejercicio 2 de las sugerencias de estudio, calcule la cohesión de grupo utilizando la ecuación 32.2.
   [Respuesta: $C_o = .40$.]

4. Lea alguno de los siguientes artículos, los cuales aplican la sociometría. Ponga especial atención a la forma en que se realizó.

   Ray, G. E., Cohen, R., Secrist, M. E. y Duncan, M. K. (1997). Relating aggressive and victimization behaviors to children's sociometric status and friendships. *Journal of Social and Personal Relationships, 14*, 95-108. [El estudio se enfoca en la relación entre nominaciones de compañeros estudiantes de 9 a 12 años de edad, sobre comportamiento agresivo y de víctima, y el estatus sociométrico: popular, promedio, rechazado y el número de amigos mutuos.]

   Schwendinger, H. y Schwendinger, J. R. (1997). Charting subcultures at a frontier of knowledge. *British Journal of Sociology, 48*, 71-94. [El artículo describe un programa de investigación para estudiar las subculturas adolescentes por medio del uso de gráficas de grandes redes subculturales. Tales gráficas y redes se generan por métodos de tipo social y sociométricos para lograr entender fenómenos como la adolescencia y la delincuencia.]

5. Consulte uno de los siguientes estudios sobre muestreo de eventos y muestreo de tiempo. Tome nota sobre la forma en que los autores ejecutan cada uno de los procedimientos.

   Bass, R. F. y Aserlind, L. (1984). Interval and time-sample data collection procedures. Methodological issues. *Advances in Learning and Behavioral Disabilities, 3*, 1-39. [Muestreo de tiempo.]

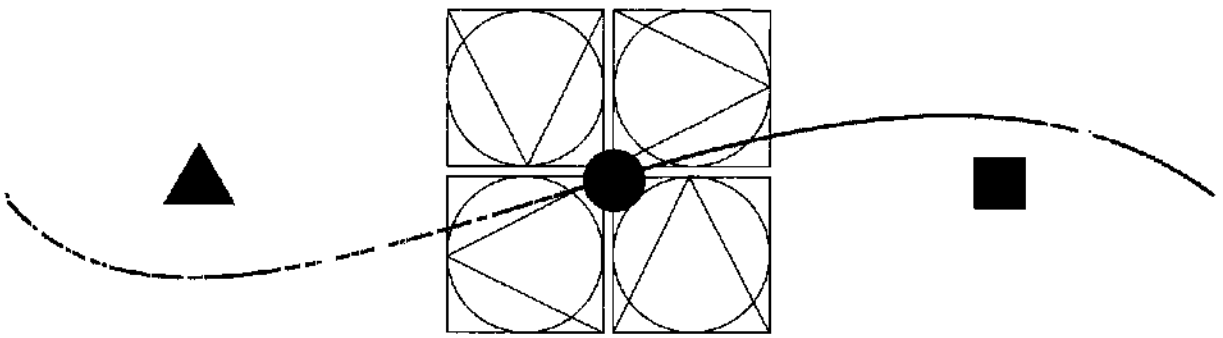
   Brown, K. W. y Moskowitz, D. S. (1998). Dynamic stability of behavior: The rhythms of our interpersonal lives. *Journal of Personality, 66*, 105-134. [Muestreo de eventos.]

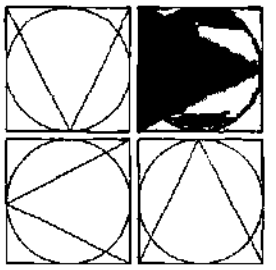
Child, G. H. (1997). A concurrent validity study of teacher's rating for nominated "problem" children. *British Journal of Educational Psychology, 67*, 457-474. [Muestreo de tiempo.]

Peregrine, P. N., Drews, D. R., North, M. y Slupe, A. (1993). Sampling techniques and sampling error in naturalistic observation: An empirical evaluation with implications for cross-cultural research. *Cross-cultural Research: Journal of Comparative Social Science, 27*, 232-246. [Compara el muestreo de eventos, de tiempo y por racimos.]

Parte Diez

Métodos multivariados





CAPÍTULO 32

ANÁLISIS DE REGRESIÓN MÚLTIPLE: FUNDAMENTOS

El *análisis de regresión múltiple* es un método para estudiar los efectos y magnitudes de los efectos de más de una variable independiente sobre una variable dependiente, utilizando los principios de correlación y regresión. Se pasará inmediatamente a la investigación y se dejará la explicación para más adelante.

Tres ejemplos de investigación

¿De qué manera se relacionan la contaminación del aire y el nivel socioeconómico con la mortalidad por padecimientos respiratorios? Lave y Seskin (1970), en su estudio de datos

de personas inglesas y estadounidenses, utilizaron el análisis de regresión múltiple para responder esa pregunta. En sus estudios realizados en barrios ingleses evaluaron los presuntos efectos de la contaminación del aire y el nivel socioeconómico, como variables independientes, sobre las tasas de mortalidad por cáncer pulmonar, bronquitis y neumonía, como variables dependientes.

El efecto general de las dos variables independientes sobre la variable dependiente se expresa por el cuadrado del coeficiente de correlación múltiple, o R^2. La interpretación de este coeficiente es similar a la de r^2, que se estudió con anterioridad. Recuerde que al elevar al cuadrado un coeficiente de correlación se produce un estimado de la cantidad de varianza compartida por dos variables. Dicho concepto es muy utilizado en el análisis de regresión.

Es la proporción de la varianza de la variable dependiente, en este caso la mortalidad, explicada por las dos variables independientes. Las R^2 entre la mortalidad debida a la bronquitis, por una parte, y la contaminación del aire y el nivel socioeconómico, por la otra, oscilaron entre .30 y .78 en diferentes muestras en Inglaterra y Gales, indicando así relaciones sustanciales. Las R^2 para las variables dependientes, mortalidades por cáncer pulmonar y por neumonía, fueron similares. La regresión múltiple también le permite al investigador aprender algo sobre las influencias relativas de las variables independientes. En la mayoría de las muestras, la contaminación del aire fue más importante que el nivel socioeconómico. Como análisis de "control", Lave y Seskin (1970) estudiaron otros tipos de cáncer que se suponía que no se verían afectados por la contaminación del aire. Las R^2 fueron consistentemente más bajas, tal como se esperaba. Extensiones de la investigación hacia áreas metropolitanas de Estados Unidos produjeron resultados similares. Los estudiantes deben tener en cuenta la explicación previa sobre la dificultad para interpretar resultados no experimentales. No obstante, Lave y Seskin construyeron un caso fuerte, a pesar de que algunas de sus interpretaciones resultaron cuestionables. Hasta este punto, sería adecuado que los lectores regresaran a la sección "Relaciones multivariadas y regresión" en el capítulo 5.

¿De qué manera se relacionan la moderación en la dieta, la ingesta energética y la actividad física con el peso corporal? Klesges, Isbell y Klesges (1992) utilizaron un análisis de regresión múltiple para intentar responder esa pregunta. El estudio es interesante a causa del número de problemas de salud asociados con la obesidad. Estos investigadores recolectaron los datos de 287 adultos, durante un año. La variable dependiente era el cambio de peso desde la línea base hasta el seguimiento un año después. Las variables independientes eran el peso en la línea base, el índice de la masa corporal, la puntuación de moderación, la edad, la ingesta energética total, el porcentaje de ingesta de grasa, el porcentaje de carbohidratos y los niveles de actividad física. Las mediciones de la estatura y del peso de tales participantes se utilizaron para estimar la masa corporal. Cada semana se tomaban las medidas del consumo de la dieta; la actividad física se midió a través de un cuestionario de actividad física con 16 reactivos que representaban actividades físicas. La restricción se midió con una escala de moderación. Se realizaron dos análisis separados de regresión; uno para los hombres y otro para las mujeres.

La R^2 entre la variable dependiente, el cambio en el peso, y una combinación lineal de las variables dependientes fue de .13 para los hombres y de .21 para las mujeres, lo cual indica que las variables independientes estudiadas por Klesges *et al.* explicaron sólo el 13 por ciento de la variabilidad observada en el cambio de peso de los hombres, y sólo el 21 por ciento en el caso de las mujeres. La cifra de los hombres no fue estadísticamente significativa al nivel $\alpha = .05$. No obstante, la ecuación de regresión para las mujeres fue significativa al nivel $\alpha = .01$. Sin embargo, a través del empleo de la regresión, estos investigadores fueron capaces de determinar que para los hombres el peso corporal y la

masa corporal iniciales representaron las dos variables más importantes para explicar el cambio en el peso.

En un estudio sobre la predicción del promedio de calificaciones de preparatoria, Holtzman y Brown (1968) utilizaron dos medidas de las variables independientes: hábitos y actitudes de estudio (*hábitos*) y la aptitud escolar (*aptitud*). Las correlaciones entre el promedio de preparatoria y los hábitos y la aptitud en primero de secundaria ($N = 1\,684$) fueron .55 y .61. La correlación entre hábitos y aptitud fue de .32. ¿Qué cantidad más de varianza se explicaría al añadir la medida de aptitud escolar a la medida de hábitos de estudio? Si se combinan hábitos y aptitud de manera óptima para predecir el promedio, se obtiene una correlación de .72. Entonces, la respuesta a la pregunta es $.72^2 - .55^2 = .52 - .30 = .22$, o 22 por ciento más de la varianza del promedio se explica al añadir la aptitud y los hábitos.

Se trata de ejemplos de análisis de regresión múltiple. La idea básica es la misma que en la correlación simple, excepto que se utilizan k variables independientes, cuando k es mayor que 1, para predecir la variable dependiente. En el análisis de regresión simple una variable, X, se utiliza para predecir otra variable, Y. En el análisis de regresión múltiple, las variables $X_1, X_2, \ldots, X_k$, sirven para predecir Y. El método y los cálculos se realizan de tal manera que den la "mejor" predicción posible, dadas las correlaciones entre todas las variables. En otras palabras, en lugar de decir "Si X entonces Y", se dice "Si $X_1, X_2, \ldots, X_k$, entonces Y", y los resultados de los cálculos indican qué tan "buena" es la predicción, y aproximadamente qué cantidad de la varianza de Y está explicada por la "mejor" combinación lineal de las variables independientes.

Análisis de regresión simple

Se afirma que se estudia la regresión de las puntuaciones de Y sobre las puntuaciones de X. Se busca estudiar de qué forma las puntuaciones de Y "regresan hacia", cómo "dependen de", las puntuaciones de X. Galton (véase Cowles, 1989), quien fue el primero en mencionar la noción de correlación, obtuvo la idea a partir del concepto de "regresión hacia la mediocridad", un fenómeno observado en estudios sobre la herencia. (El símbolo r empleado para el coeficiente de correlación, originalmente significó "regresión".) Los hombres altos tenderán a tener hijos más bajos; y los hombres bajos, hijos más altos. Entonces, las estaturas de los hijos tienden a "regresar a" o "a volver a" la media poblacional. Estadísticamente, si se desea predecir Y a partir de X, y la correlación entre X y Y es cero, entonces la mejor predicción es la media. Es decir, para cualquier X, por ejemplo X_7, sólo es posible predecir la media de Y. Sin embargo, a mayor correlación, habrá una mejor predicción. Si $r = 1.00$, entonces la predicción es perfecta. En la medida en que la correlación se aleje de 1.00, en esa misma medida las predicciones de X a Y serán menos perfectas. Si se grafican los valores de X y Y cuando $r = 1.00$, todos se encontrarán en una línea recta. A mayor correlación, más cerca estarán los valores graficados a la línea de regresión (véase capítulo 5).

Para ilustrar y explicar el concepto de regresión estadística se utilizan dos ejemplos ficticios con números simples. Los números usados en los ejemplos son iguales, excepto que se ordenan de manera diferente. Los ejemplos se toman del capítulo 15 donde, al considerar el análisis de varianza, se estudiaron los efectos de la prueba F de la correlación entre grupos experimentales. Los ejemplos se presentan en la tabla 32.1. En el ejemplo de la izquierda, denominado A, la correlación entre los valores de X y de Y es .90; mientras que en el ejemplo de la derecha, denominado B, la correlación es 0. En la tabla también se presentan ciertos cálculos necesarios para realizar el análisis de regresión: las sumas y las

TABLA 32.1 *Análisis de regresión de dos conjuntos de puntuaciones*

A. $r = .90$					B. $r = .00$				
Y	X	XY	Y'	d	Y	X	XY	Y'	d
1	2	2	1.2	−.2	1	5	5	3	−2
2	4	8	3.0	−1.0	2	2	4	3	−1
3	3	9	2.1	.9	3	4	12	3	0
4	5	20	3.9	.1	4	6	24	3	1
5	6	30	4.8	.2	5	3	15	3	2
Σ:	15	20	69	0	15	20	60		0
M:	3	4	$\Sigma d^2 = 1.90$		3	4	$\Sigma d^2 = 10.00$		
Σ^2:	55	90			55	90			

$$\Sigma y^2 = 55 - \frac{(15)^2}{5} = 10 \qquad \Sigma y^2 = 55 - \frac{(15)^2}{5} = 10$$

$$\Sigma x^2 = 90 - \frac{(20)^2}{5} = 10 \qquad \Sigma x^2 = 90 - \frac{(20)^2}{5} = 10$$

$$\Sigma xy = 69 - \frac{(15)(20)}{5} = 9 \qquad \Sigma xy = 60 - \frac{(15)(20)}{5} = 0$$

$$b = \frac{\Sigma xy}{\Sigma x^2} = \frac{9}{10} = .90 \qquad b = \frac{0}{10} = 0$$

$$a = \bar{Y} - b\bar{X} = 3 - (.90)(4) = -.60 \qquad a = 3 - (0)(4) = 3$$

$$Y' = a + bX = -.60 + .90X \qquad Y' = 3 + (0)X$$

medias, las sumas de cuadrados de la desviación de X y Y ($\Sigma x^2 = \Sigma X^2 - (\Sigma X)^2/n$), los productos cruzados de desviación ($\Sigma xy = \Sigma XY - (\Sigma X)(\Sigma Y)/n$), y ciertos valores de regresión que se explicarán en breve.

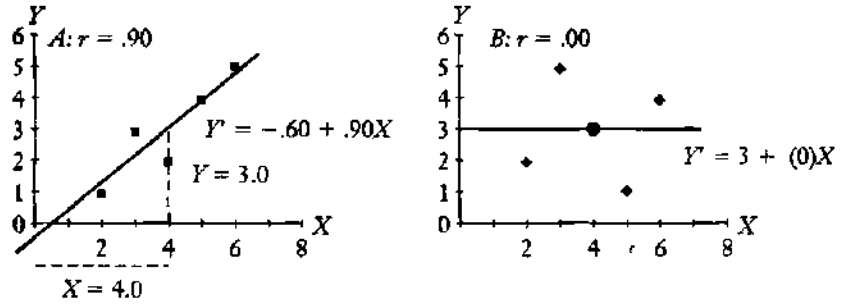
En primer lugar, observe la diferencia entre las puntuaciones en los conjuntos A y B. Difieren únicamente en el orden de las puntuaciones de la segunda columna o columna de las X. Los dos órdenes diferentes producen correlaciones muy diferentes entre las puntuaciones de X y Y. En el conjunto A, $r = .90$, y en el conjunto B, $r = .00$. En segundo lugar, observe los estadísticos en la parte inferior de la tabla. Σx^2 y Σy^2 son iguales tanto en A como en B, pero Σxy es 9 en A y 0 en B. Concentrándose en las puntuaciones del conjunto A, la ecuación básica de la regresión lineal simple es:

$$Y' = a + bX \tag{32.1}$$

donde X = las puntuaciones de la variable independiente, a = la constante de la intersección, b = el coeficiente de regresión y Y' = las puntuaciones predichas de la variable dependiente. Una ecuación de regresión es una fórmula de predicción: los valores de Y se predicen a partir de los valores de X. La correlación entre los valores observados de X y Y, en efecto, determina cómo "funciona" la ecuación de predicción. La constante de la intersección, a, y el coeficiente de regresión, b, se explicarán en breve.

Los dos conjuntos de valores X y Y de la tabla 32.1 se grafican en la figura 32.1. Para cada gráfica se dibujaron líneas que "corren a través" de los puntos graficados. Si existiera una forma de colocar dichas líneas de tal manera que estuvieran simultáneamente lo más

FIGURA 32.1



cerca posible de todos los puntos, entonces las líneas deberían expresar la regresión de Y sobre X. La línea en la gráfica de la izquierda, donde $r = .90$, corre cerca de los puntos XY graficados. No obstante, en la gráfica de la derecha, donde $r = .00$, no es posible trazar la línea cerca de todos los puntos. En efecto, los puntos están ubicados aleatoriamente, puesto que $r = .00$.

Las correlaciones entre X y Y, $r = .90$ y $r = .00$, determinan las pendientes de las líneas de regresión (cuando las desviaciones estándar de X y Y son iguales, como sucede en este caso). La *pendiente* indica el cambio que se espera en Y con el cambio de una unidad de X. En el ejemplo de $r = .90$, cuando hay un cambio de 1 en X, se predice un cambio de .90 en Y, lo cual se expresa de manera trigonométrica como la longitud de la línea opuesta al ángulo formado por la línea de regresión, dividida entre la longitud de la línea adyacente al ángulo. En la figura 32.1, si se traza una línea perpendicular a la línea de regresión —el punto donde las medias de X y de Y intersectan, por ejemplo— hasta una línea trazada horizontalmente, a partir del punto donde la línea de regresión se intersecta con el eje Y, o en $Y = -.60$, entonces $3.6/4.0 = .90$. Un cambio de 1 en X significa un cambio de .90 en Y. Se han utilizado puntuaciones en bruto en la mayor parte de los ejemplos del presente capítulo, debido a que se ajustan mejor a los propósitos del texto. Sin embargo, un examen profundo de la regresión requiere de una explicación que utilice puntuaciones de desviación y puntuaciones estándar. El énfasis aquí, así como en cualquier parte del libro, está en los usos de investigación de los métodos y técnicas, y no en la estadística como tal. Por lo tanto, el estudiante debe complementar su estudio con buenas explicaciones básicas sobre la regresión simple y múltiple. Se recomienda remitirse a las referencias de la sección de "Sugerencias de estudio", al final del capítulo 33.

La gráfica de los valores de X y Y del ejemplo B (parte derecha de la figura 32.1) es bastante diferente. En el ejemplo A, es posible, fácil y visualmente, trazar una línea a través de los puntos y lograr una aproximación bastante precisa hacia la línea de regresión. Pero en el ejemplo B esto es difícilmente posible. La línea tan sólo se puede trazar por medio del uso de otras guías, las cuales se explicarán en breve. Otro aspecto que es importante destacar es el esparcimiento o dispersión de los puntos graficados alrededor de las dos líneas de regresión. En el ejemplo A, se presentan muy cerca de la línea. Si $r = 1.00$, entonces todos estarían sobre la línea. Por otra parte, cuando $r = .00$, se dispersan ampliamente alrededor de la línea. ***A menor correlación, habrá mayor dispersión.***

Para calcular los estadísticos de regresión de los dos ejemplos, se deben calcular las sumas de cuadrados y los productos cruzados de desviación, lo cual se efectuó en la parte inferior de la tabla 32.1. La fórmula para calcular la *pendiente* o *coeficiente de regresión*, b, es:

$$b = \frac{\Sigma xy}{\Sigma x^2} \tag{32.2}$$

Las dos b son .90 y .00. La *constante de intersección*, a, se calcula por medio de la fórmula:

$$a = \bar{Y} - b\bar{X} \tag{32.3}$$

Las a para los dos ejemplos son –.60 y 3; para el ejemplo A, $a = 3 - (.90)(4) = -.60$. La constante de intersección es el punto donde la línea de regresión se intersecta con el eje Y. Para trazar la línea de regresión, se coloca una regla entre la constante de intersección en el eje Y y el punto donde se unen la media de Y y la media de X. (En la figura 32.1 dichos puntos se indican con pequeños cuadrados en el ejemplo A y con diamantes en el ejemplo B.)

Los pasos finales del proceso son, al menos hasta donde se llegará aquí, escribir ecuaciones de regresión y, después, utilizando las ecuaciones, calcular los valores predichos de Y o Y', dados los valores de X. Las dos ecuaciones se presentan en la última línea de la tabla 32.1. Primero observe la ecuación de regresión para $r = .00$: $Y' = 3 + (0)X$. Evidentemente ello significa que todas las Y predichas son iguales a 3, la media de Y. Cuando $r = 0$, la mejor predicción es la media, como ya se indicó antes. Cuando $r = 1.00$, en el otro extremo, el lector notará que es posible realizar una predicción exacta: simplemente se suma a, la constante, a las puntuaciones de X. Cuando $r = .90$, la predicción es menos que perfecta y se predicen los valores de Y' calculados con la ecuación de regresión. Por ejemplo, para predecir la primera puntuación Y', se calcula:

$$Y'_1 = -.60 + (.90)(2) = 1.20$$

Las puntuaciones predichas de los conjuntos A y B ya se presentaron en la tabla 32.1. (Véase las columnas denominadas Y'.) Note un aspecto importante: si para el ejemplo A se grafica la X y la Y predicha o los valores de Y', todos los puntos graficados caen en la línea de regresión. Es decir, la línea de regresión de la figura representa el conjunto de valores predichos de Y, dados los valores observados de X y la correlación entre la X y los valores de Y.

Ahora se pueden calcular los valores predichos de Y. A mayor correlación, habrá mayor precisión en la predicción. La precisión de las predicciones de los dos conjuntos de puntuaciones se demuestra con claridad al calcular las diferencias entre los valores originales de Y y los valores predichos de Y, o $Y - Y' = d$, y calculando, después, las sumas de cuadrados de tales diferencias, que se consideran *residuales*. En la tabla 32.1 se calcularon los dos conjuntos de residuales y sus sumas de cuadrados (véase las columnas denominadas d). Los dos valores de Σd^2, 1.90 para A y 10.00 para B, son bastante diferentes, así como las gráficas en la figura 32.1 son bastante diferentes: el valor del conjunto B, $r = .00$, es mucho mayor que el del conjunto A, o $r = .90$. Es decir, a mayor correlación, menores serán las desviaciones de la predicción y, por lo tanto, la predicción se vuelve más precisa.

En la próxima sección se examinará una extensión del modelo más simple de regresión. El modelo desarrollado, llamado regresión múltiple, posee una utilidad mucho mayor para la investigación que el modelo simple presentado aquí. Sin embargo, no se debe pensar que el modelo más simple carece de utilidad o de valor, ya que puede, de hecho, proporcionar información valiosa de investigación y de tipo práctico. Por ejemplo, Erlich y Lee (1978) demostraron cómo el modelo de regresión simple puede utilizarse con algunos estadísticos adicionales (banda o intervalo de confianza), para determinar la responsabilidad de los consejos y políticas educativas.

Regresión lineal múltiple

El método de la regresión lineal múltiple extiende las ideas presentadas en la sección anterior a más de una variable independiente. A partir del conocimiento de los valores de dos o más variables independientes, X_1, X_2,..., X_k, se desea predecir una variable dependiente, Y. Anteriormente en este libro se mencionó la gran necesidad de evaluar la influencia de diversas variables sobre una variable dependiente. Evidentemente, es posible predecir a partir de la aptitud verbal, por ejemplo, el rendimiento en lectura, o a partir del conservadurismo, las actitudes hacia los diferentes grupos étnicos. No obstante, sería más poderoso si se pudiera predecir a partir de la aptitud verbal junto con otras variables que se sabe o se piensa que influyen en la lectura —por ejemplo, la motivación de logro y la actitud hacia el trabajo escolar—. En teoría no existe límite alguno para el número de variables que se pueden utilizar, aunque existen límites prácticos. A pesar de que en el siguiente ejemplo se utilizan únicamente dos variables independientes, los principios se aplican de la misma forma a cualquier número de variables independientes.

Un ejemplo

Tome uno de los problemas que acaban de mencionarse. Suponga que se tienen las puntuaciones de rendimiento en lectura (RL), de aptitud verbal (AV) y de motivación de

TABLA 32.2 *Ejemplo ficticio: puntuación de rendimiento en lectura (Y), aptitud verbal (X_1) y motivación de logro (X_2)*

Y	X_1	X_2	Y'	$Y - Y' = d$
2	2	4	3.0305	−1.0305
1	2	4	3.0305	−2.0305
1	1	4	2.3534	−1.3534
1	1	3	1.9600	−.9600
5	3	6	4.4944	.5056
4	4	6	5.1715	−1.1715
7	5	3	4.6684	2.3316
6	5	4	5.0618	.9382
7	7	3	6.0226	.9774
8	6	3	5.3455	2.6545
3	4	5	4.7781	−1.7781
3	3	5	4.1010	−1.1010
6	6	9	7.7059	−1.7059
6	6	8	7.3125	−1.3125
10	8	6	7.8799	2.1201
9	9	7	8.9504	.0496
6	10	5	8.8407	−2.8407
6	9	5	8.1636	−2.1636
9	4	7	5.5649	3.4351
10	4	7	5.5649	4.4351
Σ: 110	99	104		0
M: 5.50	4.95	5.20		
Σ^2: 770.0	625.0	600.0		81.6091

logro (ML), de 20 alumnos de segundo año de secundaria. Se desea predecir el rendimiento en lectura, Y, a partir de la aptitud verbal, X_1, y la motivación de logro, X_2. O se desea calcular la regresión del rendimiento en lectura en *ambas* variables, tanto en la aptitud verbal como en la motivación de logro. Si las puntuaciones de aptitud verbal y motivación de logro fueran puntuaciones estandarizadas, podrían promediarse, tratar los promedios como una variable independiente compuesta y calcular los estadísticos de regresión como se realizó antes. No podría resultar muy incorrecto, sin embargo, existe una mejor forma.

Suponga que X_1 (aptitud verbal), X_2 (motivación de logro) y Y (rendimiento en lectura), las puntuaciones de los 20 sujetos y sus sumas, medias y sumas de cuadrados de los datos en bruto aparecen en la tabla 32.2. (Por el momento haga caso omiso de las columnas de Y' y de d.) Es necesario calcular las sumas de cuadrados de *desviación*, los productos cruzados de desviación, las desviaciones estándar y las correlaciones entre las tres variables, pues éstos son los estadísticos básicos que se calculan para casi cualquier conjunto de datos. (Se presentan en la tabla 32.3.) Los cálculos no se hacen aquí debido a que su mecánica se explicó en capítulos previos. El lector debe realizarlos y notar que los resultados obtenidos probablemente serán ligeramente diferentes de los reportados antes. Tales diferencias se deben a errores de redondeo: un problema siempre presente en el análisis multivariado. De hecho, los resultados de este problema, obtenidos con una calculadora de escritorio, difieren ligeramente de los obtenidos por computadora. Las sumas de cuadrados y los productos cruzados se presentan en la diagonal (de la parte superior izquierda a la parte inferior derecha) y encima de ella, y las correlaciones se presentan por debajo de la diagonal. Las r de interés primordial son las de las dos variables independientes con la variable dependiente, r_{y1} y r_{y2}, .6735 y .3946, respectivamente. Con la eliminación de estos cálculos de rutina, ahora es posible concentrarse en los conceptos básicos de la regresión múltiple. La ecuación fundamental de la regresión es:

$$Y' = a + b_1X_1 + \ldots + b_kX_k \tag{32.4}$$

Los símbolos tienen el mismo significado que los de la ecuación de regresión simple, con la excepción de que hay k variables independientes y k coeficientes de regresión. De cualquier manera, la a y las b deben calcularse a partir del conocimiento de las X y Y, cuyos cálculos son los más complejos del análisis de regresión múltiple. Para sólo dos variables independientes se utilizan las fórmulas algebraicas incluidas en los libros de estadística (véase Cohen y Cohen, 1983; Draper y Smith, 1981; Neter, Wasserman y Kutner, 1983; Pedhazur, 1996). Una vez que se tienen las b, el cálculo de a es directo. El problema es el

TABLA 32.3 *Sumas de cuadrados y productos cruzados de desviación, coeficientes de correlación y desviaciones estándar (datos de la tabla 34.2)*[a]

	y	x_1	x_2
y	165.00	100.50	39.00
x_1	*.6735*	134.95	23.20
x_2	*.3946*	*.2596*	**59.20**
s	2.9469	2.6651	1.7652

[a] Los datos de la tabla son los siguientes: la primera línea da, de forma sucesiva, Σy^2, la suma de cuadrados de las puntuaciones de desviación para Y, el producto cruzado de las desviaciones de X_1 y Y, o Σx_1y y finalmente Σx_2y. Las cifras en la segunda y tercera líneas, sobre la diagonal o encima de ella, son Σx_1^2, Σx_1x_2, y (en la esquina de la parte baja derecha) Σx_2^2. Las cifras en *itálicas* (*cursivas*) por *debajo* de la diagonal son los coeficientes de correlación. Las desviaciones estándar se presentan en la última línea.

cálculo de las b cuando hay más de dos variables independientes. Sólo se explicarán las ideas generales que subyacen a los cálculos, ya que los detalles nos desviarían del tema de interés central. Se le pide al lector que consulte cualquiera de las referencias que se mencionan arriba.

Lo que se tiene, en efecto, es un conjunto de ecuaciones lineales, una ecuación para cada variable independiente. El objetivo de la determinación de las b de la ecuación 32.4 consiste en encontrar aquellos valores de b que minimicen las sumas de cuadrados de los residuales. Éste es el *principio de los mínimos cuadrados*. El cálculo proporciona el método de diferenciación para llevarlo a cabo. Si se utiliza, produce un conjunto de ecuaciones lineales simultáneas llamadas ecuaciones *normales* (que no tienen relación con las distribuciones normales). Una forma conveniente de estas ecuaciones contiene los coeficientes de correlación entre todas las variables independientes, y entre las variables independientes y la variable dependiente y un conjunto de ponderaciones llamadas *pesos beta*, β_j, que se explicarán posteriormente (son como los pesos b). Las ecuaciones normales para el problema anterior son:

$$\begin{aligned} r_{11}\beta_1 + r_{12}\beta_2 &= r_{y1} \\ r_{12}\beta_1 + r_{22}\beta_2 &= r_{y2} \end{aligned} \tag{32.5}$$

donde β_j es igual a los pesos beta; r_{12} es igual a las correlaciones entre las variables independientes, y r_{yj} a las correlaciones entre las variables independientes y la variable dependiente, Y. (Observe que $r_{12} = r_{21}$, y que $r_{11} = r_{22} = 1.00$, y también que la ecuación 32.5 puede extenderse para cualquier número de variables independientes.)

Quizá la mejor manera —y con seguridad la más elegante— de resolver las ecuaciones para β_j sea por medio del álgebra de matrices. Por desgracia, el conocimiento del álgebra de matrices no puede darse como un hecho. Por lo tanto, la solución real de las ecuaciones, utilizando el álgebra de matrices, debe omitirse; pero la solución para las dos variables independientes se obtiene algebraicamente sin el uso de matrices. Sin embargo, cualquier tamaño mayor que éste seguramente requeriría el empleo de un programa computacional, ya que la cantidad de cálculos se incrementa de manera exponencial. Se mostrará cómo se obtiene la solución con un poco de álgebra. Para utilizar la ecuación normal dada anteriormente, se necesitará calcular r_{12}, r_{y1} y r_{y2}. Recuerde que tanto r_{11} como r_{22} son iguales a 1.00. A partir de examinar la tabla 32.3, se obtiene $r_{12} = .259562$, $r_{y1} = .6735$ y $r_{y2} = .394604$. Sustituyendo estos valores en la ecuación normal, se obtiene:

$$1.000000\beta_1 + .259562\beta_2 = .6735$$

$$.259562\beta_1 + 1.00000\beta_2 = .394604$$

Tome cada ecuación normal y despéjela para que β_1 esté a un lado del signo de igual y el resto quede del otro lado:

$$\beta_1 = .6735 - .259562\beta_2$$

$$\beta_1 = 1.5202688 - 3.8526441\beta_2$$

Ahora iguale las β_1 entre sí y despéjelas para obtener β_2.

$$.6735 - .259562\beta_2 = 1.5202688 - 3.852641\beta_2$$

$$3.5930821\beta_2 = .8467688$$

$$\beta_2 = \frac{.8467688}{3.5930821} = .235666 \approx .2357$$

Despejando para β_1 se sustituye el valor de β_2 en la ecuación: $\beta_1 = .6735 - .259562\beta_2$. Por lo tanto,

$$\beta_1 = .6735 - .259562 \times .235666 = .6123$$

La solución para la situación de dos variables presentada arriba produce los siguientes pesos beta: $\beta_1 = .6123$ y $\beta_2 = .2357$. Los pesos b o los pesos no estandarizados de la regresión se obtienen a partir de la siguiente fórmula:

$$b_j = \beta_j \frac{s_y}{s_j} \tag{32.6}$$

donde s_j es igual a las desviaciones estándar de las variables uno y dos (véase tabla 32.3) y s_y es igual a la desviación estándar de Y. Sustituyendo en la ecuación 32.6 se obtiene:

$$b_1 = (.6123)\left(\frac{2.9469}{2.6651}\right) = .6771$$

$$b_2 = (.2357)\left(\frac{2.9469}{1.7652}\right) = .3934$$

Para obtener la constante de la intersección, se extiende la ecuación 32.3 a dos variables independientes:

$$a = \bar{Y} - b_1\bar{X}_1 - b_2\bar{X}_2$$

$$a = 5.50 - (.6771)(4.95) - (.3934)(5.20) = .1027$$

Un método alternativo para encontrar los coeficientes no estandarizados de la regresión es por medio de ecuaciones normales. Note aquí que dichas ecuaciones normales darán directamente los pesos de regresión. Las ecuaciones normales dadas arriba serían utilizadas para obtener los coeficientes de regresión.

$$nb_0 + \Sigma x_1 b_1 + \Sigma x_2 b_2 = \Sigma y$$

$$\Sigma x_1 b_0 + \Sigma x_1^2 b_1 + \Sigma x_1 x_2 b_2 = \Sigma x_1 y$$

$$\Sigma x_2 b_0 + \Sigma x_1 x_2 b_1 + \Sigma x_2^2 b_2 = \Sigma x_2 y$$

Observe que para dos variables independientes hay tres ecuaciones normales y que algunas veces se utiliza b_0 para representar el término o constante de la intersección. Se tienen tres ecuaciones arriba con tres parámetros desconocidos para estimarse: b_0, b_1 y b_2. Los cálculos para resolver los pesos de la regresión son demasiado laboriosos para realizarse a mano y, por lo tanto, no se presentarán los cálculos aquí.

Por último, se escribe la ecuación de regresión completa:

$$Y' = a + b_1X_1 + b_2X_2$$

$$Y' = .1027 + .6771X_1 + .3934X_2$$

Sustituyendo los valores observados de X_1 y X_2 de la tabla 34.2, se obtienen los valores predichos de Y. Por ejemplo, calcule las Y' predichas para el quinto y vigésimo sujetos:

$$Y'_5 = .1027 + (.6771)(3) + (.3934)(6) = 4.4944$$

$$Y'_{20} = .1027 + (.6771)(4) + (.3934)(7) = 5.5649$$

Estos valores y los otros 18 se presentan en la cuarta columna de la tabla 32.2. La quinta columna de la tabla presenta las desviaciones a partir de la regresión, o los residuales, $Y_i - Y'_i = d_i$. Por ejemplo, los residuales para Y_5 y Y_{20} son:

$$d_5 = Y_5 - Y'_5 = 5 - 4.4944 = .5056$$

$$d_{20} = Y_{20} - Y'_{20} = 10 - 5.5649 = 4.4351$$

Observe que una desviación es pequeña y la otra es grande. Los residuales se presentan en la última columna de la tabla 32.2. La mayoría de ellos son relativamente pequeños; casi la mitad son positivos y la otra mitad negativos.

Ahora puede calcularse la suma de cuadrados debida a la regresión; sin embargo, debe considerarse la regresión de Y sobre X_1 y X_2. Se eleva al cuadrado cada uno de los valores de Y' de la cuarta columna de la tabla 32.2 y se suman:

$$(3.0305)^2 + \cdots + (5.5649)^2 = 688.3969$$

Ahora, se utiliza la fórmula acostumbrada para la suma de cuadrados de desviación (véase capítulo 13):

$$\sum y^2 = 688.3969 - \frac{(110)^2}{20} = 83.3969$$

De manera similar, calcule la suma de cuadrados de los residuales:

$$\sum d^2 = (-1.0305)^2 + \cdots + (4.4351)^2 = 81.6091$$

Note que éste es un "buen" ejemplo de los errores que se acumulan debido al redondeo. La verdadera suma de cuadrados de la regresión, calculada por medio de una computadora, es 83.3909, un error de .006. No obstante, también debe notarse que aunque los residuales se calcularon a partir de las Y predichas calculadas a mano, la suma de cuadrados de los residuales es exactamente igual a la producida por medio de la computadora, 81.6091.

Para verificar, calcule:

$$SC_{reg} + SC_{res} = SC_t$$

$$83.3969 + 81.6091 = 165.0060$$

La regresión y las sumas de cuadrados residuales por lo común no se calculan de esta manera. Aquí se calcularon tan sólo para mostrar cuáles son estas cantidades. Si se hubieran utilizado las fórmulas empleadas comúnmente, entonces quizá no se habría visto con claridad que la suma de cuadrados de regresión es la suma de cuadrados de los valores de Y', calculados por medio de la ecuación de regresión. Quizá tampoco se hubiera visto con claridad que la suma de cuadrados residual es la suma de cuadrados calculada con las d de la quinta columna de la tabla 32.2. Recuerde también que la a y las b (o las β) de la ecuación de regresión se calcularon para satisfacer el principio de los mínimos cuadrados; es decir,

para minimizar las d o errores de predicción o, más bien, para minimizar la suma de cuadrados de los errores de predicción. Para sintetizar, la suma de cuadrados de regresión expresa aquella porción de la suma de cuadrados total de Y debida a la regresión de Y, la variable dependiente, sobre X_1 y X_2, las variables independientes. La suma de cuadrados residual expresa la porción de la suma de cuadrados total de Y que *no* se debe a la regresión.

Tal vez el lector se pregunte: ¿Para qué tomarse la molestia con este procedimiento complicado para determinar los pesos de regresión? ¿Es necesario recurrir a un procedimiento de mínimos cuadrados? ¿Por qué no sólo promediar los valores de X_1 y X_2, y llamar a las medias de los valores individuales de X_1 y X_2 las Y predichas? La respuesta es que funcionaría bastante bien. De hecho, en este caso funcionaría muy bien, casi tan bien como el procedimiento total de regresión. Pero quizá *no* funcionaría tan bien. El problema radica en que realmente no se sabe cuándo funcionará bien y cuándo no. El procedimiento de regresión generalmente "funciona", siempre y cuando las cosas se mantengan iguales. Siempre minimiza los errores cuadrados de predicción. Note que en ambos casos se utilizan ecuaciones lineales y que sólo difieren los coeficientes:

$$\textit{Ecuación de regresión: } Y' = a + b_1X_1 + b_2X_2$$

$$\textit{Ecuación media: } Y' = \frac{1}{2}X_1 + \frac{1}{2}X_2$$

De las innumerables formas posibles para ponderar X_1 y X_2, ¿cuál debe elegirse si no se utiliza el principio de los mínimos cuadrados? Por supuesto, es concebible que se tenga conocimiento previo o alguna razón para X_1 y X_2. Las X_1 pueden ser las puntuaciones en alguna prueba que haya demostrado ser altamente exitosa a nivel predictivo. X_2 puede ser un predictor exitoso también, pero no tanto como X_1. De esta manera, se toma la decisión de darle un gran peso a X_1, por ejemplo, cuatro veces más que a X_2. La ecuación sería: $Y' = 4X_1 + X_2$; lo cual podría funcionar bien. El problema es que en pocas ocasiones se tiene este conocimiento previo, y aun cuando se tenga, suele resultar bastante impreciso. ¿Cómo se puede llegar a decidir darle cuatro veces más peso a X_1 que a X_2? Es posible efectuar una suposición con cierto fundamento. No obstante, el método de regresión no es una suposición; se trata de un método preciso basado en los datos y en un principio matemático poderoso. Es en tal sentido que los pesos calculados de regresión son "mejores".

Las sumas de cuadrados de regresión y residuales se calculan de forma más directa de lo indicado antes. Las fórmulas son:

$$sc_{reg} = b_1\Sigma x_1y + \ldots + b_k\Sigma x_ky \tag{32.7}$$

$$sc_{res} = sc_t - sc_{reg} \tag{32.8}$$

En el presente caso, (32.7) se convierte en:

$$sc_{reg} = b_1\Sigma x_1y + b_2\Sigma x_2y$$

Esto se calcula fácilmente sustituyendo los dos valores de b calculados arriba y los productos cruzados dados en la tabla 32.3.

$$sc_{reg} = (0.6771)(100.50) + (0.3934)(39.00) = 83.3912$$

$$sc_{res} = 165.0 - 83.3912 = 81.6088$$

Con los errores por redondeo, éstos son los valores calculados de forma directa, a partir de la cuarta y quinta columnas de la tabla 32.2. (Note los valores "más precisos" dados por

una computadora: $sc_{reg_M} = 83.3909$ y $sc_{res} = 81.6091$, que evidentemente da un total $sc_t = \Sigma y^2 = 165.0$.)

El coeficiente de correlación múltiple

Si se calcula el coeficiente común de correlación producto-momento entre los valores predichos Y', y los valores observados de Y, se obtiene un índice de la magnitud de la relación entre un compuesto de los mínimos cuadrados de X_1 y X_2, por un lado, y Y, por el otro. Este índice se llama *coeficiente de correlación múltiple*, R. A pesar de que en el presente capítulo casi siempre se simboliza como R por cuestiones de brevedad, una manera más satisfactoria de hacerlo es con subíndices: $R_{y.12...k}$ o, en este caso, $R_{y.12}$. La teoría de la regresión múltiple parece ser especialmente elegante cuando se considera el coeficiente de correlación múltiple. Se trata de uno de los vínculos que une los diversos aspectos de la regresión múltiple y del análisis de varianza. La fórmula de R que expresa el primer enunciado de este párrafo es:

$$R = \frac{\Sigma yy'}{\sqrt{\Sigma y^2 \Sigma y'^2}} \tag{32.9}$$

Se calcula su cuadrado:

$$R^2 = \frac{(\Sigma yy')^2}{\Sigma y^2 \Sigma y'^2} \tag{32.10}$$

Utilizando los valores de Y y Y' de la tabla 32.2, se obtiene: $R^2 = .5054$ y $R = \sqrt{.5054} = .7109$. El cálculo de estos valores es un buen ejercicio. Ya se tiene $\Sigma y^2 = 165$. Entonces se calcula:

$$\Sigma y'^2 = \Sigma Y'^2 - \frac{(\Sigma Y')^2}{N} = 688.3969 - \frac{(110)^2}{20} = 83.3969$$

y

$$\Sigma yy' = \Sigma YY' - \frac{(\Sigma Y)(\Sigma Y')}{N} = 688.3939 - \frac{(110)(110)}{20} = 83.3939$$

Puede demostrarse algebraicamente que $\Sigma y'^2$ es igual a $\Sigma yy'$. La diferencia de .003 se debe a los errores de redondeo.

Entonces, R es la correlación más alta posible entre un compuesto lineal de los mínimos cuadrados de las variables independientes y la variable dependiente observada. La R^2, análoga a r^2, indica la porción de varianza de la variable dependiente, Y, debida a las variables independientes en conjunto. R, a diferencia de r, varía únicamente de 0 a 1.00; no posee valores negativos.

Otras dos conclusiones importantes se obtienen calculando las correlaciones de los *residuales, d, de la tabla 32.2, con X_1 y X_2, por un lado, y con Y, por el otro. Las correlaciones de los residuales con X_1 y X_2 son ambas cero, lo cual* no sorprende cuando se sabe que,

por definición, los residuales son aquella parte de Y que no está explicada por X_1 y X_2. Es decir, cuando los valores de Y' se restan de los valores de Y, aquella porción debida a la regresión de Y sobre X_1 y X_2 se toma de ellos. Cualquier cosa que quede, entonces, no se relaciona con X_1 ni con X_2. Si el lector se tomara la molestia de calcular la correlación entre el vector de d —un vector es un solo conjunto de medidas, ya sea en una columna o en un renglón— y el vector de X_1 o de X_2 se podrá observar que esto es cierto. No se debe subestimar la importancia de realizar dichos cálculos y de ponderar su significado. Ello es especialmente importante como ayuda para comprender la regresión múltiple y otras técnicas multivariadas. Puede cometerse un serio error al permitir que la computadora haga todo por nosotros, especialmente con programas computacionales. En lo que se refiere a los estadísticos más sencillos como r y las diversas sumas de cuadrados, es mejor escribir programas relativamente simples para una microcomputadora, almacenarlos en discos flexibles y utilizarlos cuando así se requiera. Una importante implicación para la investigación de esta generalización también se analizará posteriormente cuando se sinteticen y expongan ejemplos reales de investigación.

La correlación de los residuales, d_i, de la tabla 32.2 con los valores originales de Y, también ayuda a aclarar algunos puntos. Esta correlación es: $r_{dy} = .7033$, y su cuadrado es: $r_{dy}^2 = (.7033)^2 = .4946$. Si este último valor se suma a la R^2 calculada anteriormente, el resultado es interesante: $R^2 + r_{dy}^2 = .5054 + .4946 = 1.0000$; lo cual siempre será cierto: "1.0000" representa la varianza total de Y. La varianza de Y debida a la regresión de las Y sobre X_1 y X_2 es .5054. Se puede calcular la varianza de Y que no se debe a la regresión de Y sobre X_1 y X_2: $1.0000 - .5054 = .4946$ que es, evidentemente, el valor de r_{dy}^2 que acaba de calcularse de manera directa. El significado de r_{dy}^2 se interpreta de dos formas. El cálculo directo de la correlación muestra que los residuales constituyen la parte de la varianza de Y que no se debe a la regresión de Y a partir de X_1 y X_2. En el presente caso, el 51 por ciento ($R^2 = .51$) de la varianza del rendimiento en lectura (Y) de los 20 alumnos se explica por una combinación lineal de los mínimos cuadrados de la aptitud verbal (X_1) y la motivación de logro (X_2). Pero el 49 por ciento de la varianza se debe a otras variables y al error. Después de analizar formas más comunes para calcular R y R^2, se considerará nuevamente la interpretación de la proporción o porcentaje de R^2.

En síntesis, R^2 es un estimado de la proporción de la varianza de la variable dependiente Y, explicado por las variables independientes X_j. R, el coeficiente de correlación múltiple, es la correlación producto-momento entre la variable dependiente y otra variable producida por una combinación de los mínimos cuadrados de las variables independientes. Su cuadrado se interpreta de manera análoga al cuadrado de un coeficiente de correlación ordinario. Sin embargo, difiere del coeficiente ordinario en que únicamente toma valores entre 0 y 1. La R no es tan útil ni tan interpretable como la R^2, y de aquí en adelante se utilizará la R^2 casi de manera exclusiva en los análisis presentados.

La interpretación de la proporción o del porcentaje de R^2 se vuelve más clara si se utiliza una fórmula de suma de cuadrados:

$$R^2 = \frac{sc_{reg}}{sc_t} \tag{32.11}$$

donde sc_t es, como siempre, la suma de cuadrados total de Y, o $\sum_{yt}^2$. Sustituyendo la suma de cuadrados de regresión calculada antes por medio de la fórmula 32.7, y la suma de cuadrados total de la tabla 32.3, se obtiene:

$$R^2 = \frac{83.3912}{165.000} = .5054$$

Y R^2 parece ser esa parte de la suma de cuadrados de Y asociada con la regresión de Y a partir de las variables independientes. Como sucede con todas las proporciones, al multiplicarla por 100 se convierte en un porcentaje.

La fórmula 32.11 proporciona otro vínculo con el análisis de varianza. En el capítulo 13, al explicar los fundamentos del análisis de varianza, se presentó una fórmula para calcular η, la *razón de correlación* (fórmula 13.4). Se eleva al cuadrado dicha fórmula:

$$\eta^2 = \frac{sc_e}{sc_t}$$

dónde sc_e es igual a la suma de cuadrados entre grupos, y sc_t es igual a la suma de cuadrados total. sc_e es la suma de cuadrados debida a la variable independiente. sc_{reg} es la suma de cuadrados debida a la regresión. Ambos términos se refieren a la suma de cuadrados de una variable dependiente, debida a una variable independiente o a variables independientes.

R y R^2 quizá estén infladas, lo que sucede con frecuencia. Por lo tanto, R^2 debe interpretarse de manera conservadora. Si la muestra es grande, por ejemplo, mayor de 200, entonces existen pocas razones para preocuparse; sin embargo, si la muestra es pequeña, resulta sensato reducir algunos puntos a la R^2 calculada. Para hacerlo, se utiliza una *fórmula de encogimiento*:

$$R_c^2 = 1 - (1 - R^2)\left(\frac{N-1}{N-n-1}\right)$$

donde R_c^2 es igual a la R^2 encogida o corregida; N es el tamaño de la muestra; n es el número total de variables en el análisis. Usando esta fórmula, la R^2 en el ejemplo se reduce a .45. Cuando se compara R_c^2 con R^2, se percibe qué tanto está inflada la R^2 por el error del azar. A partir de dicha fórmula también se observa el efecto que tiene un tamaño pequeño de muestra sobre el valor de R_c^2. Las muestras pequeñas tienden a producir valores inestables de R^2, lo cual se determina por medio de la fórmula de encogimiento antes expresada.

Pruebas de significancia estadística

Anteriormente se estudió la regresión simple de Y a partir de X. Para probar la significancia estadística de la regresión simple se evalúa la significancia del coeficiente de correlación entre X y Y, r_{xy}, refiriéndose a una tabla apropiada. Algunos de los libros reconocidos que contienen tablas utilizadas en análisis estadísticos son los de Beyer (1990) y Burlington y May (1970). Con el avance en los programas estadísticos computacionales y su fácil acceso, los investigadores utilizan cada vez menos las revisiones de tablas. Los programas computacionales ahora son capaces de calcular y dar el resultado de la probabilidad del error tipo I, junto con el estadístico de prueba, haciendo innecesaria la consulta de las tablas. Sin embargo, desde un punto de vista educativo, los estudiantes necesitan aprender el manejo de las tablas para comprender el resultado que ofrece la computadora. Las pruebas de significancia estadística en la regresión múltiple, aunque son más complejas, se basan en la idea relativamente simple de la comparación de varianzas (o cuadrados medios), como en el análisis de varianza. Las mismas preguntas planteadas muchas veces antes, deben plantearse de nuevo: ¿puede esta R^2 haber surgido por azar? ¿Se aleja lo suficiente de lo esperado por el azar como para considerarla "significativa"? Preguntas similares pueden plantearse acerca de los coeficientes individuales de regresión. En este capítulo y en el siguiente, se utilizarán casi exclusivamente las pruebas F, pues se ajustan bastante bien tanto al análisis de regresión como al análisis de varianza y, además, son simples tanto

a nivel conceptual como a nivel computacional. Es posible realizar análisis sobre cada coeficiente de regresión. Si la prueba t de un coeficiente de regresión es significativa, ello indica que el peso de regresión difiere significativamente de cero, lo cual, a su vez, significa que la variable con la que está asociado contribuye de manera significativa a la regresión. La prueba t para coeficientes de regresión individuales se presenta en la siguiente sección. Primero se utilizará la prueba F para determinar si el modelo completo de regresión resulta estadísticamente significativo.

Una forma se expresa por medio de las ecuaciones 32.12a y 32.12b

$$F = \frac{sc_{reg}/gl_1}{sc_{res}/gl_2} \tag{32.12a}$$

$$F = \frac{sc_{reg}/k}{sc_{res}/(N-k-1)} \tag{32.12b}$$

donde sc_{reg} es la suma de cuadrados debida a la regresión; sc_{res} es igual a la suma de cuadrados residual o de error; k es igual al número de variables independientes; N es igual al tamaño de la muestra. Si se definen gl_1 y gl_2, los grados de libertad para el numerador y el denominador de la razón F, en la ecuación 32.12a, se obtiene la ecuación 32.12b. Tales fórmulas son importantes a causa de que se utilizan para probar la significancia de cualquier problema de regresión múltiple. Con los valores calculados antes para el ejemplo de la tabla 32.2, ahora se calcula:

$$F = \frac{83.3912/2}{81.6091/(20-2-1)} = \frac{41.6956}{4.8005} = 8.686$$

Note que la idea expresada por esta fórmula pertenece a la misma familia de ideas que el análisis de varianza. El numerador es el cuadrado medio debido a la regresión, análogo al cuadrado medio entre grupos, y el denominador es el cuadrado medio que *no* se debe a la regresión, el cual se utiliza como un término de error, análogo al cuadrado medio dentro de grupos o varianza del error. Nuevamente, el principio básico es siempre el mismo: la varianza debida a la regresión de Y a partir de $X_1, X_2, \ldots, X_k$, o, en el análisis de varianza, debida a los efectos experimentales, se evalúa en contra de la varianza debida presuntamente al error o al azar. Dicho concepto básico, elaborado con profundidad en capítulos previos, se expresa de la siguiente forma:

$$\frac{\text{varianza de regresión}}{\text{varianza de error}} : \frac{\text{varianza experimental}}{\text{varianza de error}}$$

Otra fórmula para F es:

$$F = \frac{R^2/k}{(1-R^2)/(N-k-1)} \tag{32.13}$$

donde k y N son iguales que arriba. Para el mismo ejemplo:

$$F = \frac{.5054/2}{(1-.5054)/(20-2-1)} = \frac{.2527}{.0291} = 8.684$$

el cual es igual al valor de F obtenido con la ecuación 32.12, incluyendo los errores debidos al redondeo. Con 2 y 17 grados de libertad, es significativo al nivel .01. Esta fórmula resulta particularmente útil cuando los propios datos de investigación aparecen sólo en forma de coeficientes de correlación. En tal caso, quizá no se conozcan las sumas de cuadrados requeridas por la ecuación 32.12. Gran parte del análisis de regresión puede realizarse usando únicamente la matriz de correlaciones entre todas las variables, independientes y dependientes. Dicho análisis va más allá del enfoque de este libro. No obstante, el lector y el estudiante de investigación deben estar conscientes de esa posibilidad (Pedhazur, 1996).

Pruebas de significancia de los coeficientes individuales de regresión

La significancia de los pesos o coeficientes individuales de regresión interesa a muchos investigadores, pues les indican cuáles variables independientes, en un sentido estadístico, realizan la principal contribución para explicar la variable dependiente. Por ejemplo, el encargado de admisiones de una importante universidad tiene a su disposición un número de variables que son pertinentes para predecir o explicar el éxito en la universidad. No obstante, con un análisis real, algunas de estas variables tendrían una contribución considerablemente menor que otras. Por ejemplo, McWhirter (1997) fue capaz de identificar qué variables predecían la soledad íntima y la soledad social de estudiantes universitarios.

La fórmula para probar la significancia de los pesos o coeficientes individuales de regresión es:

$$t_i = \frac{b_i}{s_{bi}}$$

donde b_i es el coeficiente de regresión y s_{bi} es el error estándar de la variable i. La fórmula anterior parece lo suficientemente simple; sin embargo, el cálculo del error estándar resulta complejo. La mejor manera para obtener el error estándar es por medio de un programa computacional o, en caso necesario, del álgebra matricial. Esta prueba t se conduce con grados de libertad igual a n – (el número de coeficientes de regresión en la ecuación de regresión).

Interpretación de los estadísticos de regresión múltiple

La interpretación de los estadísticos de la regresión múltiple llega a ser compleja y difícil. De hecho, la interpretación de los estadísticos del análisis multivariado es, en general, considerablemente más difícil que la interpretación de los estadísticos univariados estudiados previamente. Por lo tanto, se profundizará en cierto grado en la interpretación de los estadísticos del ejemplo.

Significancia estadística de la regresión y de R^2

La razón F de 8.684, calculada antes, indica que la regresión de Y sobre X_1 y X_2, expresada por $R^2_{y.12}$, es estadísticamente significativa. La probabilidad de que una razón F tan grande

como ésta, ocurra debido al azar, es menor al .01 (en realidad es aproximadamente .003), lo cual quiere decir que la relación entre Y y una combinación de los mínimos cuadrados de X_1 y X_2 quizá no ocurrió debido al azar.

La $R = .71$ se interpreta de forma similar a un coeficiente de correlación ordinario, excepto que los valores de R oscilan entre 0 y 1.00, a diferencia de la r, que va de –1.00 a 1.00, pasando por cero. Sin embargo, $R^2 = .71^2 = .51$ es más útil y tiene mayor importancia: significa que el 51 por ciento de la varianza de Y se explica o "determina" por X_1 y X_2, en combinación. Se le denomina, en concordancia, *coeficiente de determinación*. Una denominación alternativa para este estadístico es *CMC*, que son las siglas de "correlación múltiple cuadrada".

Contribuciones relativas de X a Y

Permítase el planteamiento, un tanto tímido, de una pregunta más difícil: ¿cuáles son las contribuciones relativas de X_1 y de X_2, de la aptitud verbal y la motivación de logro, a Y, el rendimiento en lectura? El enfoque restringido de este libro no permite el examen que merecen las respuestas a esta pregunta. El problema de la contribución relativa de las variables independientes a una variable o variables dependientes es uno de los más complejos y difíciles de los análisis de regresión. Parece que en realidad no existe ninguna solución satisfactoria, al menos cuando se correlacionan las variables independientes. No obstante, el problema no puede soslayarse. Sin embargo, el lector debe tener en cuenta que hay que tener mucha reserva en relación con la explicación anterior y las posteriores. Los problemas técnicos y sustantivos de la interpretación del análisis de regresión múltiple se tratan en dos o tres de las referencias presentadas en el ejercicio 1 de la sección de sugerencias de estudio, en el capítulo 33.

Se pensaría que los coeficientes de regresión, b o β, proporcionarían medios preparados para identificar las contribuciones relativas de las variables independientes a una variable dependiente; y así es, pero sólo de forma aproximada y en algunas ocasiones de forma confusa. Anteriormente se mencionó que el coeficiente de regresión b se llama *pendiente*. La pendiente de la línea de regresión está en relación con el porcentaje de unidades b de Y para una unidad de X. En el problema A de la tabla 32.1, por ejemplo, $b = .90$. Así, como se indicó antes, con el cambio de una unidad en X se predice un cambio de .90 en Y. No obstante, en la regresión múltiple, una interpretación tan directa como ésta no es tan fácil, debido a que existe más de una b. Sin embargo, se puede decir, *para los propósitos pedagógicos presentes*, que si X_1 y X_2 tienen aproximadamente la misma escala de valores —en el ejemplo de la tabla 32.2 los valores de X_1 y X_2 están en el rango aproximado del 1 al 10— las b son pesos que muestran un aproximado de la importancia relativa de X_1 y X_2. En el presente caso, la fórmula de regresión es:

$$Y' = .1027 + .6771X_1 + .3934X_2$$

Se puede decir que X_1, la aptitud verbal, tiene mayor peso que X_2, la motivación de logro, lo cual es verdad para este caso, pero quizá no siempre sea así, especialmente con más variables independientes.

Los coeficientes de regresión, por desgracia para los propósitos de interpretación, no permanecen estables, pues cambian con diferentes muestras y con la suma o resta de variables independientes en el análisis (véase Dillon y Goldstein, 1984; Howell, 1997; Pedhazur, 1996). No existe una forma absoluta para interpretarlos. Si todas las correlaciones entre las variables independientes son iguales o cercanas a cero, entonces se simplifica mucho la interpretación. Pero muchas, o la mayoría, de las variables que se correlacionan con la

TABLA 32.4 *Ejemplos de regresión múltiple con y sin correlaciones entre las variables independientes*

A			B		
1	2	*Y*	1	2	*Y*
1.00	.50	.87	1.00	0	.87
.50	1.00	.43	0	1.00	.43
.87	.43	1.00	.87	.43	1.00
$R^2_{y.12} = .76$			$R^2_{y.12} = .94$		

variable dependiente también se correlacionan entre sí. El ejemplo de la tabla 32.3 indica lo siguiente: la correlación entre X_1 y X_2 es .26, una correlación modesta. Sin embargo, dichas intercorrelaciones con frecuencia son más altas; y cuanto más altas sean (hasta cierto punto), más inestable será la situación de interpretación.

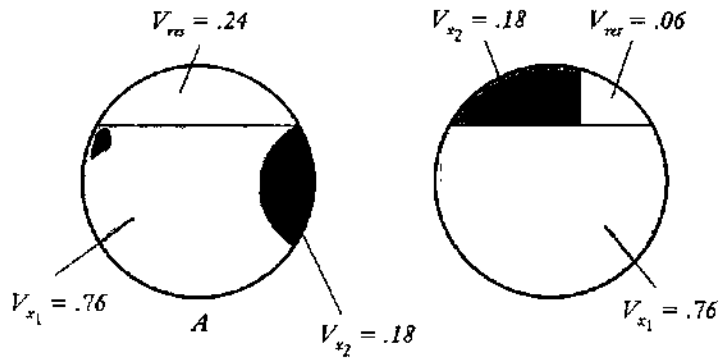
La situación predictiva ideal ocurre cuando las correlaciones entre las variables independientes y la variable dependiente son altas, y cuando las correlaciones entre las variables independientes son bajas. Este principio es importante. A mayor intercorrelación entre las variables independientes, habrá mayor dificultad en la interpretación. Entre otras cuestiones, se tiene una gran dificultad para establecer la influencia relativa de las variables independientes sobre la variable dependiente. Examine las dos matrices de correlación ficticias de la tabla 32.4 y sus R^2 acompañantes. En las dos matrices las variables independientes, X_1 y X_2, tienen una correlación de .87 y .43, respectivamente, con la variable dependiente, Y. Pero las correlaciones entre las variables independientes difieren en los dos casos. En la matriz A, $r_{12} = .50$, una correlación importante. En la matriz B, sin embargo, $r_{12} = 0$.

El contraste entre las R^2 es notorio: .76 para A y .94 para B. Puesto que en B, X_1 y X_2 no están correlacionadas, entonces, cualesquiera correlaciones que tengan con Y contribuyen de forma directa a la predicción y a la R^2. Cuando las correlaciones entre las variables independientes son exactamente iguales a cero, como en la matriz B, entonces resulta fácil calcular la R^2; simplemente es la suma de cuadrados de las r, entre cada variable independiente y la variable dependiente: $(.87)^2 + (.43)^2 = .94$. Cuando las variables independientes están correlacionadas, como sucede en la matriz A ($r_{12} = .50$), una parte de la varianza común de Y y X_1 también es compartida por X_2. En síntesis, X_1 y X_2 son redundantes, hasta cierto punto, al predecir Y. En la matriz B no existe dicha redundancia.

La situación se aclara, quizás, por medio de la figura 32.2. Sean los círculos la varianza total de Y, y que dicha varianza total sea 1.00. Entonces, las porciones de la varianza de Y, explicadas por X_1 y X_2, pueden representarse. En ambos círculos, el sombreado gris claro indica la varianza explicada para X_1 o V_{X1}; y el sombreado gris oscuro, para X_2 o V_{X2}. (Las varianzas restantes después de V_{X1} y V_{X2} son las varianzas residuales, denominadas así en la figura.) En B, V_{X1} y V_{X2} no se sobreponen. Sin embargo, en A, V_{X1} y V_{X2} sí se sobreponen. Sencillamente, debido a que $r_{12} = 0$ en B y $r_{12} = .50$ en A, el poder predictivo de las variables independientes es mucho mayor en B que en A. Por supuesto, ello se ve reflejado por las R^2: .76 en A y .94 en B.

Aunque se trata de un ejemplo ficticio y artificial, tiene la virtud de mostrar el efecto de correlación entre las variables independientes y, por lo tanto, ilustra el principio enunciado antes. También refleja la dificultad para interpretar los resultados de la mayoría de los análisis de regresión, ya que en muchas investigaciones las variables independientes

FIGURA 32.2



$R^2_{y \cdot 12} = .76$
$R^2_{y \cdot 12} = V_{x_1} + V_{x_2}$
$= .76 + 0 = .76$

$R^2_{y \cdot 12} = .94$
$R^2_{y \cdot 12} = V_{x_1} + V_{x_2}$
$= .76 + .18 = .94$

están correlacionadas. Y cuando se añaden más variables independientes, la interpretación se torna aún más compleja y difícil. Un problema central es: ¿cómo se clasifican los efectos relativos de las diferentes X sobre Y? La respuesta también es compleja. Existen varias formas de hacerlo, algunas más satisfactorias que otras; aunque ninguna lo es completamente. Quizás la forma más satisfactoria, por lo menos según la opinión y experiencia de los autores, se logra calculando *correlaciones semiparciales cuadradas* (también llamadas *correlaciones por partes*), las cuales se calculan con la fórmula:

$$SP^2 = R^2_{y.12\ldots k} - R^2_{y.12\ldots(k-1)}$$

o en el presente caso, para B:

$$SP^2 = R^2_{y.12} - R^2_{y.1} = .94 - .76 = .18$$

que indica la contribución de X_2 a la varianza de Y, *después de que* X_1 *se ha tomado en cuenta.* El mismo cálculo para A produce: .76 – .76 = 0, lo cual indica que X_2 no contribuye en lo absoluto a la varianza de Y, después de que X_1 se ha tomado en cuenta. (En realidad existe un ligero incremento que surge únicamente con un gran número de decimales.)

Se refiere al lector a Howell (1997), Dillon y Goldstein (1984) y Pedhazur (1996) para un análisis sobre los problemas involucrados. Kerlinger y Pedhazur (1973) también tratan el problema con considerable detalle y lo relacionan con ejemplos de investigación.

Otros problemas analíticos y de interpretación

Una cantidad de problemas del análisis de regresión múltiple no se tratan en este libro con el detalle que merecen. Sin embargo, es necesario mencionar algunos de ellos a causa de la creciente importancia de la regresión múltiple en la investigación del comportamiento.

Uno de ellos, ya mencionado, es el problema de los pesos de la regresión. En este capítulo y en el siguiente, la exposición se centra en los pesos b, ya que en la mayor parte de los usos de investigación de la regresión se predice con las puntuaciones por renglón o de desviación, y las b se utilizan con dichas puntuaciones. Beta o los pesos β, por otro lado, se utilizan con las puntuaciones estándar. Se les llama *coeficientes de regresión parcial estandarizados*. "Estandarizados" significa que serían utilizados si todas las variables estuvieran en forma de puntuación estándar. "Parcial" significa que los efectos de variables, distintas de aquella donde se aplica el peso, se mantienen constantes. Por ejemplo, $\beta_{y.123}$ o β_1, en un problema de tres variables (independientes), es el peso estandarizado de la regresión parcial, el cual expresa el cambio en Y, debido al cambio en X_1, manteniendo constantes las variables dos y tres. Un segundo significado, utilizado en el trabajo teórico, es que b es el peso de regresión poblacional que β estima. Aquí se omite ese significado. Las β se traducen en b con la fórmula:

$$b_j = \beta_j \frac{s_y}{s_j}$$

donde s_y es igual a la desviación estándar de Y y s_j la desviación estándar de la variable j. Los pesos b también son coeficientes parciales de regresión, pero no están en forma estandarizada.

Otro problema es que en cualquier regresión dada, R, R^2 y los pesos de la regresión serán iguales, sin importar el orden de las variables. No obstante, si se suma o se resta una o más variables de la regresión, dichos valores cambiarán. Y los pesos de regresión pueden cambiar de muestra a muestra. En otras palabras, no existe una calidad absoluta respecto a ellos. Por ejemplo, no puede decirse que debido a que las aptitudes verbal y numérica tienen pesos de regresión de .60 y .50 en un conjunto de datos, tendrán los mismos valores en otro conjunto.

Anteriormente en este libro se dijo: "El diseño es la disciplina de los datos." El diseño de investigación y el análisis de datos surgen a partir de las demandas de los problemas de investigación. De nuevo, el orden de aparición de las variables independientes en la ecuación de regresión se determina por el problema de investigación y el diseño de la investigación, el cual, a su vez, se determina por el problema de investigación.

Aunque el orden de anotación de las variables y los cambios en los pesos de regresión que ocurren con muestras que difieren son problemas difíciles, es necesario recordar que los pesos de regresión finales no cambian con órdenes diferentes de anotación. Ésta es una verdadera compensación, especialmente útil en la predicción. Por ejemplo, en muchos problemas de investigación, la contribución relativa de las variables no es una consideración importante. En tales casos, la ecuación total de la regresión y sus pesos de regresión se requieren principalmente para predecir y para evaluar la naturaleza general de la situación de regresión.

Sin embargo, cuando el investigador desea encontrar la contribución de cada variable independiente, deben utilizarse los pesos beta (pesos de regresión estandarizados). Dichos pesos beta se han escalado, de tal manera que se comparan entre sí de manera directa. Los pesos de regresión no estandarizados reflejan la escala de medición utilizada para medir esa variable. Por lo tanto, los pesos de regresión no estandarizados no pueden compararse de forma directa. Además, la significancia de los pesos beta es igual a la significancia del cambio en R^2, cuando una variable independiente se anota al final, dentro de la ecuación de regresión.

Otro aspecto importante es que, por lo general, existe una utilidad limitada en la añadidura de variables a una ecuación de regresión. Debido a que muchas variables de

investigación del comportamiento se correlacionan, opera el principio ilustrado por los datos de la tabla 32.4 y analizado anteriormente, disminuyendo la utilidad de variables adicionales. Si se encuentran tres o cuatro variables independientes que estén altamente correlacionadas con una variable dependiente, y no altamente correlacionadas entre sí, entonces se tiene suerte. Pero se vuelve cada vez más difícil encontrar otras variables dependientes que no sean, en efecto, redundantes con las primeras tres o cuatro. Si $R^2_{y.123} = .50$, entonces es poco probable que $R^2_{y.1234}$ sea mucho mayor que .55, y $R^2_{y.12345}$ tal vez no será mayor que .56 o .57. Se tiene una ley de regresión de ganancias reducidas. Cuando se agregan variables independientes, se nota cuánto agregan a R^2 y se prueba su significancia estadística. La fórmula para hacerlo, similar a la fórmula 32.13, es:

$$F = \frac{(R^2_{y.12.k_1} - R^2_{y.12.k_2})/(k_1 - k_2)}{(1 - R^2_{y.12.k_1})/(N - k_1 - 1)}$$

donde k_1 es el número de variables independientes de la R^2 mayor, k_2 es el número de variables independientes de la R^2 menor, y N es igual al número de casos. Tal fórmula se utilizará posteriormente. A pesar de que una F calculada como ésta puede ser estadísticamente significativa, en especial con una muestra grande, el incremento real de R^2 llega a ser muy pequeño. En un estudio realizado por Layton y Swanson (1958), la añadidura de una sexta variable independiente produjo una razón F estadísticamente significativa, ¡pero el incremento real de R^2 fue de .0147! La diferencia entre las R^2 en el numerador es el coeficiente de correlación semiparcial elevado al cuadrado.

Anteriormente se dijo que R, R^2 y los coeficientes de regresión permanecen iguales si las mismas variables se anotan en diferente orden. Sin embargo, no debe pensarse que ello significa que no importa el orden en que se anotan las variables en la ecuación de regresión. Por el contrario, el orden de anotación suele ser muy importante. Cuando las variables independientes están correlacionadas, la cantidad relativa de varianza de la variable dependiente explicada o a la que contribuye cada variable independiente, cambia drásticamente con un orden diferente de anotación de las variables. Con los datos A de la tabla 32.4, por ejemplo, si se revierte el orden de X_1 y X_2, sus contribuciones relativas cambian marcadamente. Con el orden original, X_2 no contribuye en nada a R^2; mientras que con el orden revertido, X_2 se convierte en X_1 y contribuye en un 19 por ciento [$r^2 = (.43)^2 = .19$] a la R^2 total, y la X_1 original, que se convierte en X_2, contribuye un 57 por ciento (.19 + .57 = .76). El orden de las variables, aunque no produce ninguna diferencia en la R^2 final y, por lo tanto, en la predicción general, constituye un problema importante en investigación.

Sin embargo, la multicolinealidad o las variables independientes correlacionadas no siempre son indeseables. En algunos casos, cuando se utiliza la regresión múltiple para establecer la validez de una medida o escala, las variables independientes correlacionadas son de gran utilidad. Las variables independientes que tienen correlaciones de cero o cercanas a cero con la variable dependiente, pero una alta correlación con otra variable independiente, en realidad llegan a incrementar la cantidad de varianza compartida por las variables dependiente e independiente. Este tipo de variable independiente se llama *variable supresora*. Algunos investigadores, como el doctor Leonard Helmers de Nueva Orleans,[1] se refieren a ellas como variables "de recorte". Dichas variables poseen el efecto de eliminar, suprimir o recortar la varianza irrelevante en las otras variables independientes. Suponga que se desea desarrollar una ecuación de regresión para predecir habilidades mecánicas. Se podría utilizar la puntuación de la persona en una prueba de desempeño de

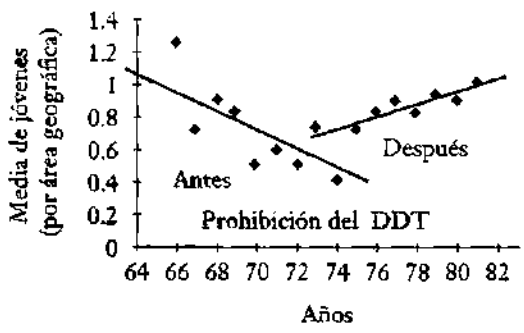
[1] Comunicación personal. El doctor Helmers fue Director de Investigación de ASI Marketing, Inc., en Hollywood, California.

habilidades mecánicas como variable dependiente, y se seleccionaría una prueba escrita sobre aptitud mecánica como variable independiente (predictora). Tal vez se desearía incluir una variable supresora, como la comprensión lectora, la cual sería un candidato para funcionar como variable supresora, ya que muy probablemente no se correlaciona con las habilidades mecánicas, pero sí con la *prueba escrita de aptitud mecánica* (Mechanical Aptitude Test), pues dicha prueba requiere de lectura. Así, las dos variables independientes *aptitud mecánica* y *comprensión lectora*, pueden correlacionarse, y la prueba de desempeño mecánico (Mechanical Performance Test) puede correlacionarse con la prueba de aptitud mecánica; sin embargo, la prueba de comprensión lectora no estaría correlacionada con la prueba de desempeño mecánico.

En otras situaciones, un investigador quizá no esté consciente de que una variable supresora se haya utilizado en el análisis. ¿Entonces cómo se puede saber si en este caso tiene una variable supresora? Bueno, si se cuenta con un programa computacional como el SPSS o el SAS (Sistema de Análisis Estadístico), el resultado generado por estos programas para computadora puede, bajo un escrutinio cuidadoso, utilizarse para detectar la presencia de una variable supresora. El primer paso consiste en determinar qué variables independientes poseen un peso beta distinto a cero. Si se encuentra una y el valor absoluto de la correlación simple entre la variable dependiente y su variable independiente es considerablemente menor que el peso beta asociado con esa variable independiente, es posible obtener una variable supresora. Además, si el peso beta para esa variable independiente es diferente a cero, y la correlación simple entre la variable dependiente y la variable independiente tiene un signo opuesto al del peso de beta, entonces ésta es una señal de que la variable independiente puede ser una variable supresora. En algunos análisis de investigación, como aquellos encontrados en la investigación de mercado, las variables supresoras se obtienen del análisis y se calcula nuevamente la ecuación de regresión.

En la literatura se encuentran ejemplos de variables supresoras. Hadfield, Littleton, Steiner y Woods (1998) utilizaron la regresión múltiple para analizar las habilidades pedagógicas de estudiantes y para probar la hipótesis de que el conocimiento de contenido matemático sería el correlato más significativo con la eficacia de la microenseñanza. [Se utilizaron las calificaciones en una filmación para medir la eficacia de la enseñanza. Las variables independientes fueron el conocimiento de contenido pedagógico, el conocimiento de contenido matemático, ansiedad ante las matemáticas y habilidad espacial.] Los autores encontraron que las puntuaciones del conocimiento de contenido matemático y de la ansiedad ante las matemáticas actuaron como variables supresoras. Con estas variables dentro de la ecuación de regresión hubo un incremento del 25 por ciento en la varianza

FIGURA 32.3



investigación del comportamiento se correlacionan, opera el principio ilustrado por los datos de la tabla 32.4 y analizado anteriormente, disminuyendo la utilidad de variables adicionales. Si se encuentran tres o cuatro variables independientes que estén altamente correlacionadas con una variable dependiente, y no altamente correlacionadas entre sí, entonces se tiene suerte. Pero se vuelve cada vez más difícil encontrar otras variables dependientes que no sean, en efecto, redundantes con las primeras tres o cuatro. Si $R^2_{y.123} = .50$, entonces es poco probable que $R^2_{y.1234}$ sea mucho mayor que .55, y $R^2_{y.12345}$ tal vez no será mayor que .56 o .57. Se tiene una ley de regresión de ganancias reducidas. Cuando se agregan variables independientes, se nota cuánto agregan a R^2 y se prueba su significancia estadística. La fórmula para hacerlo, similar a la fórmula 32.13, es:

$$F = \frac{(R^2_{y.12.k_1} - R^2_{y.12.k_2})/(k_1 - k_2)}{(1 - R^2_{y.12.k_1})/(N - k_1 - 1)}$$

donde k_1 es el número de variables independientes de la R^2 mayor, k_2 es el número de variables independientes de la R^2 menor, y N es igual al número de casos. Tal fórmula se utilizará posteriormente. A pesar de que una F calculada como ésta puede ser estadísticamente significativa, en especial con una muestra grande, el incremento real de R^2 llega a ser muy pequeño. En un estudio realizado por Layton y Swanson (1958), la añadidura de una sexta variable independiente produjo una razón F estadísticamente significativa, ¡pero el incremento real de R^2 fue de .0147! La diferencia entre las R^2 en el numerador es el coeficiente de correlación semiparcial elevado al cuadrado.

Anteriormente se dijo que R, R^2 y los coeficientes de regresión permanecen iguales si las mismas variables se anotan en diferente orden. Sin embargo, no debe pensarse que ello significa que no importa el orden en que se anotan las variables en la ecuación de regresión. Por el contrario, el orden de anotación suele ser muy importante. Cuando las variables independientes están correlacionadas, la cantidad relativa de varianza de la variable dependiente explicada o a la que contribuye cada variable independiente, cambia drásticamente con un orden diferente de anotación de las variables. Con los datos A de la tabla 32.4, por ejemplo, si se revierte el orden de X_1 y X_2, sus contribuciones relativas cambian marcadamente. Con el orden original, X_2 no contribuye en nada a R^2; mientras que con el orden revertido, X_2 se convierte en X_1 y contribuye en un 19 por ciento [$r^2 = (.43)^2 = .19$] a la R^2 total, y la X_1 original, que se convierte en X_2, contribuye un 57 por ciento (.19 + .57 = .76). El orden de las variables, aunque no produce ninguna diferencia en la R^2 final y, por lo tanto, en la predicción general, constituye un problema importante en investigación.

Sin embargo, la multicolinealidad o las variables independientes correlacionadas no siempre son indeseables. En algunos casos, cuando se utiliza la regresión múltiple para establecer la validez de una medida o escala, las variables independientes correlacionadas son de gran utilidad. Las variables independientes que tienen correlaciones de cero o cercanas a cero con la variable dependiente, pero una alta correlación con otra variable independiente, en realidad llegan a incrementar la cantidad de varianza compartida por las variables dependiente e independiente. Este tipo de variable independiente se llama *variable supresora*. Algunos investigadores, como el doctor Leonard Helmers de Nueva Orleans,[1] se refieren a ellas como variables "de recorte". Dichas variables poseen el efecto de eliminar, suprimir o recortar la varianza irrelevante en las otras variables independientes. Suponga que se desea desarrollar una ecuación de regresión para predecir habilidades mecánicas. Se podría utilizar la puntuación de la persona en una prueba de desempeño de

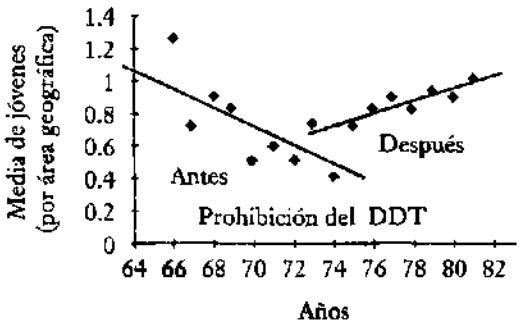
[1] Comunicación personal. El doctor Helmers fue Director de Investigación de ASI Marketing, Inc., en Hollywood, California.

habilidades mecánicas como variable dependiente, y se seleccionaría una prueba escrita sobre aptitud mecánica como variable independiente (predictora). Tal vez se desearía incluir una variable supresora, como la comprensión lectora, la cual sería un candidato para funcionar como variable supresora, ya que muy probablemente no se correlaciona con las habilidades mecánicas, pero sí con la *prueba escrita de aptitud mecánica* (Mechanical Aptitude Test), pues dicha prueba requiere de lectura. Así, las dos variables independientes *aptitud mecánica* y *comprensión lectora*, pueden correlacionarse, y la prueba de desempeño mecánico (Mechanical Performance Test) puede correlacionarse con la prueba de aptitud mecánica; sin embargo, la prueba de comprensión lectora no estaría correlacionada con la prueba de desempeño mecánico.

En otras situaciones, un investigador quizá no esté consciente de que una variable supresora se haya utilizado en el análisis. ¿Entonces cómo se puede saber si en este caso tiene una variable supresora? Bueno, si se cuenta con un programa computacional como el SPSS o el SAS (Sistema de Análisis Estadístico), el resultado generado por estos programas para computadora puede, bajo un escrutinio cuidadoso, utilizarse para detectar la presencia de una variable supresora. El primer paso consiste en determinar qué variables independientes poseen un peso beta distinto a cero. Si se encuentra una y el valor absoluto de la correlación simple entre la variable dependiente y su variable independiente es considerablemente menor que el peso beta asociado con esa variable independiente, es posible obtener una variable supresora. Además, si el peso beta para esa variable independiente es diferente a cero, y la correlación simple entre la variable dependiente y la variable independiente tiene un signo opuesto al del peso de beta, entonces ésta es una señal de que la variable independiente puede ser una variable supresora. En algunos análisis de investigación, como aquellos encontrados en la investigación de mercado, las variables supresoras se obtienen del análisis y se calcula nuevamente la ecuación de regresión.

En la literatura se encuentran ejemplos de variables supresoras. Hadfield, Littleton, Steiner y Woods (1998) utilizaron la regresión múltiple para analizar las habilidades pedagógicas de estudiantes y para probar la hipótesis de que el conocimiento de contenido matemático sería el correlato más significativo con la eficacia de la microenseñanza. [Se utilizaron las calificaciones en una filmación para medir la eficacia de la enseñanza. Las variables independientes fueron el conocimiento de contenido pedagógico, el conocimiento de contenido matemático, ansiedad ante las matemáticas y habilidad espacial.] Los autores encontraron que las puntuaciones del conocimiento de contenido matemático y de la ansiedad ante las matemáticas actuaron como variables supresoras. Con estas variables dentro de la ecuación de regresión hubo un incremento del 25 por ciento en la varianza

FIGURA 32.3



explicada. No obstante, ambas variables tuvieron una baja correlación con la variable dependiente, la eficacia de la enseñanza, pero estaban correlacionadas con las otras variables independientes.

El estudio de Leichtman y Erickson (1979) sobre los determinantes cognitivos, demográficos y de interacción de las habilidades para asumir roles en niños de cuarto grado, desarrollaron una ecuación de regresión donde cinco variables predijeron el 36 por ciento de la varianza de las puntuaciones del asumir roles. Estas variables fueron la puntuación de la escala de *vocabulario del WISC*, los errores de la prueba de emparejamiento de figuras familiares (Matching Familiar Figures Test), el sexo, el vecindario y la dominancia de mano. Se encontró que la puntuación de la escala de *vocabulario del WISC* era una variable supresora, la cual se correlacionaba con las otras variables independientes, pero no con la variable dependiente: la puntuación respecto a asumir roles.

Ejemplos de investigación

El DDT y las águilas calvas

Una de las diversas controversias sobre el deterioro del ambiente ocasionado por intereses comerciales, y la oposición y las protestas de los grupos ambientalistas se ha centrado en el uso del DDT. Uno de los efectos de rociar DDT ha sido la disminución de especies de aves. Por ejemplo, la reproducción de la población del águila calva fue seriamente afectada. En diciembre de 1972, la Agencia de Protección Ambiental (Enviromental Protection Agency) prohibió el uso del DDT. Grier (1982) en un estudio sobre el efecto que tuvo la prohibición en la reproducción de las águilas calvas, reportó el número promedio de águilas jóvenes por área geográfica, durante los años 1966 a 1981. Sus análisis de regresión (y de otros tipos) de los promedios de reproducción (medias) antes y después de la prohibición, mostraron que las dos pendientes, o coeficientes b, difirieron de forma significativa. De 1966 a 1974, $b = -.07$, lo cual indica un decremento en la reproducción a lo largo de esos años, pero de 1973 a 1981, $b = .07$, lo que indica un incremento. (Ambas b fueron estadísticamente significativas.) El método para comparar pendientes de manera estadística se presenta en Pedhazur (1996), Howell (1997) y Lee y Little (1996). Las regresiones simples se calculan utilizando los años como variable independiente y las tasas de reproducción como variable dependiente. La correlación antes de la prohibición del DDT fue de –.74, pero después de la prohibición fue de .80 (cálculos aproximados de los autores de este libro). Se graficaron dos regresiones en la figura 32.3. La gráfica refleja la regresión de la media de águilas jóvenes por área geográfica durante los años 1966 a 1974 (antes de la prohibición del DDT) y durante los años 1975-1981 (después de la prohibición). La regresión para los datos antes de la prohibición se calculó desde 1974, ya que se podía esperar que el efecto de la prohibición no se manifestara durante un año, aproximadamente. Grier realizó este cálculo desde 1973. La marcada diferencia entre las dos relaciones o pendientes es drástica.

Sesgo por exageración en exámenes de autoevaluación

¿Dicen la verdad, con frecuencia, los solicitantes de un empleo, respecto a sus capacidades? Los contratantes se preocupan cada vez más respecto al hecho de si las credenciales presentadas por un solicitante a un empleo son verdaderas. Anderson, Warner y Spencer (1984) utilizaron la regresión múltiple para responder esta pregunta. Los participantes en

TABLA 32.7 *Cantidad de varianza explicada por el examen de autoevaluación y la escala de exageración, sobre el desempeño en mecanografía*

Variable independiente	R^2	ΔR^2	
Autoevaluación	.07	.07	$p < .05$
Escala de exageración	.23	.16	$p < .05$
Interacción	.25	.02	$p > .05$

el estudio fueron 351 solicitantes de empleo para un puesto en el estado de Colorado. Se les pidió a los participantes que indicaran su nivel de experiencia con cierto tipo de tareas de trabajo. Algunas de las tareas presentadas a los solicitantes estaban diseñadas para fines de la investigación. Los investigadores crearon una escala sobre el sesgo por exageración para determinar el grado en que los solicitantes exageraban sobre sí mismos. Se utilizó un análisis de regresión múltiple para determinar la validez de esta escala de medición. A los solicitantes para puestos administrativos se les pidió que indicaran cuántas palabras por minuto podían mecanografiar, además de completar la escala de sesgo por exageración. Entonces, los investigadores usaron la prueba de mecanografía como variable dependiente en una regresión múltiple con la escala de sesgo por exageración y el examen de autoevaluación. Los resultados de la investigación se presentan en la tabla 32.7. La correlación entre la prueba de mecanografía y el examen de autoevaluación fue de .27 ($r^2 = .073$) y .41 ($r^2 = .168$) para la escala de exageración. En lo que se refiere a la explicación de la variación en las puntuaciones de la prueba de mecanografía, la escala de exageración incrementó la posibilidad de predicción. A pesar de que Anderson *et al.* no utilizaron una variable supresora para incrementar la R^2 en su estudio, podrían haberlo hecho. Un posible candidato para el trabajo de variable supresora quizás habría sido la comprensión lectora. Puesto que se requiere de la comprensión lectora para leer el cuestionario de autoevaluación, y a que probablemente esté correlacionada con el desempeño real en mecanografía, quizá habría sido una variable supresora.

Análisis de regresión múltiple e investigación científica

La regresión múltiple está cercana al corazón de la investigación científica. También es fundamental para la estadística y la inferencia, y está íntimamente relacionada con métodos matemáticos básicos y poderosos. Además, desde el punto de vista del investigador, es útil y práctica: realiza su trabajo analítico de manera exitosa y eficiente. Mediante la explicación de estas fuertes y radicales declaraciones, es posible aclarar lo que se ha aprendido.

El científico está relacionado, básicamente, con proposiciones del tipo "si p, entonces q", las cuales "explican" fenómenos. Cuando se dice "si incentivo positivo, entonces alto rendimiento", hasta cierto punto se está "explicando" el rendimiento. Pero ello no es suficiente. Aun cuando estuviera apoyada por una gran cantidad de evidencia empírica, esta proposición no va demasiado lejos en la explicación del rendimiento. Además de otras proposiciones *si-entonces* de tipo similar, el científico debe plantear preguntas más complejas. Él podría preguntar, por ejemplo, en qué condiciones es válida la proposición "si incentivo positivo, entonces alto rendimiento". ¿Es verdadera tanto para niños estadounidenses negros como para niños estadounidenses blancos? ¿Es verdadera para niños con baja y alta inteligencia? Para probar tales preguntas y para impulsar el conocimiento, los científicos escriben proposiciones del tipo *si p, entonces q, bajo las condiciones r, s y t*, donde p es una variable independiente; q una variable dependiente, y r, s y t otras variables inde-

pendientes. También se pueden escribir, evidentemente, otras proposiciones —por ejemplo, *si p y r, entonces q*—. En este caso *p* y *r* son dos variables independientes, las cuales se requieren para *q*.

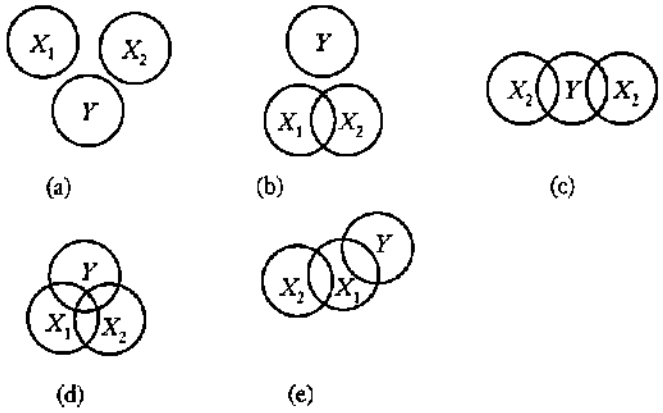
El punto central aquí es que la regresión múltiple puede manejar dichos casos de manera exitosa. En la mayor parte de la investigación del comportamiento existe, por lo general, una variable dependiente, aunque no se restringe teóricamente a una sola. Como consecuencia, la regresión múltiple constituye un método general para analizar muchos datos de investigación del comportamiento. Ciertos otros métodos se consideran casos especiales de regresión múltiple. El más relevante es el análisis de varianza, donde todos sus tipos se conceptualizan y logran con análisis de regresión múltiple.

Anteriormente se mencionó que todo control se refiere a control de la varianza. El análisis de regresión múltiple se considera como un método refinado y poderoso del "control" de varianza. Esto lo logra de la misma forma en que lo hace el análisis de varianza: estimando la magnitud de las diferentes fuentes de influencia sobre *Y*, de las diferentes fuentes de varianza de *Y*, a través del análisis de las interrelaciones de todas las variables. Indica qué cantidad de *Y* presuntamente se debe a $X_1, X_2, \ldots, X_k$. Da cierta idea sobre las cantidades relativas de influencia de las *X*; y facilita pruebas de significancia estadística de influencias combinadas de las *X* sobre *Y*, y de las influencias separadas de cada *X*. En síntesis, el análisis de regresión múltiple constituye una técnica eficiente y poderosa de comprobación de hipótesis y para realizar inferencias. Lo es debido a que ayuda a los científicos a estudiar, con relativa precisión, interrelaciones complejas entre variables independientes y una variable dependiente; *Y*, por lo tanto, los ayuda a "explicar" el presunto fenómeno representado por la variable dependiente.

RESUMEN DEL CAPÍTULO

1. La regresión múltiple constituye un método para estudiar los efectos y la magnitud de los efectos de más de una variable independiente, sobre una variable dependiente.
2. La regresión simple incluye una variable independiente y una variable dependiente.
3. A través del método de mínimos cuadrados, la regresión múltiple implica encontrar los mejores pesos de regresión que maximicen la relación entre una combinación lineal de las variables independientes y la variable dependiente.
4. *R* es la correlación múltiple. Es la correlación entre los valores reales de la variable dependiente y los valores que se predicen de la variable dependiente.
5. El cuadrado de la correlación múltiple, R^2, es un estadístico utilizado para determinar la calidad de la ecuación de regresión encontrada a través de los datos empíricos.
6. Los cálculos de la regresión múltiple son complicados. Se recomienda el uso de un programa computacional.
7. El cuadrado de la correlación múltiple, o coeficiente de determinación, se usa en pruebas estadísticas para determinar si la ecuación de regresión está explicando una cantidad significativa de la varianza.
8. Un problema de la regresión múltiple es que las variables independientes pueden estar correlacionadas. Cuando así sucede, conducen a estimados inestables de los coeficientes de regresión y a dificultades de interpretación.
9. Puede probarse la significancia estadística de la ecuación de regresión completa, así como de cada peso individual de regresión.
10. Las pruebas de los pesos individuales de regresión informarán al investigador qué variable está contribuyendo a la explicación de la variable dependiente.

FIGURA 32.4



SUGERENCIAS DE ESTUDIO

1. Lea uno o más de los siguientes estudios que utilizaron regresión múltiple. Tome nota de las variables utilizadas, de los programas computacionales utilizados y de las conclusiones obtenidas a partir de los resultados.

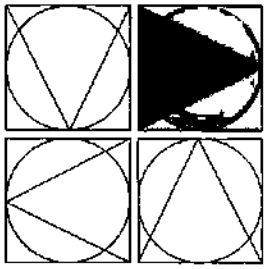
   Abel, M. H. (1998). Interaction of humor and gender in moderating relationships between stress and outcomes. *Journal of Psychology, 132*, 267-276.

   Connelly, C. D. (1998). Hopefulness, self-esteem, and perceived social support among pregnant and nonpregnant adolescents, *Western Journal of Nursing Research, 20*, 195-209.

   Ho, R. (1998). The intention to give up smoking: Disease versus social dimensions. *Journal of Social Psychology, 138*, 368-380.

   Stalenheim, E. G., Eriksson, E., von Knorring, L. y Wide, L. (1998). Testosterone as a biological marker in psychopathy and alcoholism. *Psychiatry Research*, 77, 79-88.

2. Dados los diagramas de Venn en la figura 32.4, una variable dependiente, Y, y dos variables independientes, X_1 y X_2,
   a) Determine cuál posee una variable supresora.
   b) ¿Cuál es ideal para una regresión múltiple?
   c) Indique cuál(es) exhibe(n) multicolinealidad.
   d) ¿Cuál produciría una ecuación de regresión inútil?
   e) ¿Cuál tiene mayor probabilidad de producir pruebas estadísticas no significativas en la ecuación de regresión?
   f) ¿Cuál tiene mayor probabilidad de producir pruebas estadísticas no significativas en los coeficientes individuales de regresión?



CAPÍTULO 33

REGRESIÓN MÚLTIPLE, ANÁLISIS DE VARIANZA Y OTROS MÉTODOS MULTIVARIADOS

- ANÁLISIS DE VARIANZA DE UN FACTOR Y ANÁLISIS DE REGRESIÓN MÚLTIPLE
- CODIFICACIÓN Y ANÁLISIS DE DATOS
- ANÁLISIS FACTORIAL DE VARIANZA, ANÁLISIS DE COVARIANZA Y ANÁLISIS RELACIONADOS
- ANÁLISIS DISCRIMINANTE, CORRELACIÓN CANÓNICA, ANÁLISIS MULTIVARIADO DE VARIANZA Y ANÁLISIS DE RUTA
- REGRESIÓN DE CRESTA, REGRESIÓN LOGÍSTICA Y ANÁLISIS LOGARÍTMICO LINEAL
  - **Tablas de contingencia de múltiples factores y análisis log-lineal**
- ANÁLISIS MULTIVARIADO E INVESTIGACIÓN CIENTÍFICA

Un examen detallado muestra que las bases conceptuales de los diferentes modelos de análisis de datos son iguales o similares. La simetría de las ideas fundamentales tiene un gran atractivo estético, y en ninguna parte es más interesante y atractiva que en la regresión múltiple y en el análisis de varianza. Anteriormente, cuando se explicaron los fundamentos del análisis de varianza, se resaltó la similitud entre los principios y estructuras del análisis de varianza y los denominados métodos de correlación. Ahora se vincularán los dos métodos y, en el proceso, se mostrará que el análisis de varianza puede realizarse utilizando la regresión múltiple. Además, el enlace de los dos métodos producirá felizmente ganancias inesperadas. Se observará, por ejemplo, que ciertos problemas analíticos que no pueden tratarse con el análisis de varianza —o al menos difíciles de tratar cuando no verdaderamente inapropiados— se conceptualizan y abordan con facilidad por medio del uso juicioso y flexible de la regresión múltiple y sus variantes. Por limitaciones de espacio y como el propósito del libro no es enseñar los mecanismos de los modelos ni los métodos estadísticos,

la exposición será breve: algunas de las cosas que se digan deberán creerse con base en la confianza del lector. No obstante, aun con un nivel de exposición así se verá que ciertos problemas difíciles asociados con el análisis de varianza se manejan de manera natural y fácil con el análisis de regresión múltiple. Algunos de estos problemas asociados incluyen el análisis de covarianza, los datos del pretest y del postest, inequidad del número de casos en las casillas (o diseños factoriales), variables dependientes categóricas, tablas de contingencia de múltiples factores y el manejo tanto de datos experimentales como no experimentales.

Análisis de varianza de un factor y análisis de regresión múltiple

Suponga que se llevó a cabo un experimento con tres métodos de presentación de materiales verbales a niños de tercer año de secundaria. La variable dependiente es la comprensión, medida a través de una prueba objetiva sobre los materiales. Suponga también que los resultados son los que se presentan en la tabla 33.1. Evidentemente hay que realizar un análisis de varianza. Los resultados del análisis de varianza se indican en la parte final de la tabla. Se invita al lector a que realice los cálculos de los ejemplos de este capítulo, lo cual es sumamente necesario para lograr la cabal comprensión de importantes aspectos que se tratarán aquí. Por ejemplo, realice los cálculos del análisis de varianza de la tabla 33.1 y de la regresión múltiple del problema de la tabla 33.2; estudie y pondere los resultados de ambos análisis. Es importante que no se realicen por medio de un programa computacional, ya que es probable que no se comprendan. Siempre que sea posible, debe trabajar los problemas con detenimiento. Si se sucumbe a la tentación de utilizar uno de los paquetes para las grandes computadoras o para microcomputadoras, se debe ser cauteloso, pues algunas ocasiones la calidad de los programas estadísticos para las microcomputadoras (y para las grandes) es cuestionable. La razón F en la tabla 33.1 es 18, la cual con 2 y 12 grados de libertad, es significativa al nivel .01. El efecto del tratamiento experimental es claramente significativo: $\eta^2 = sc_t/sc_t = 90/120 = .75$. La relación entre el tratamiento experimental y la comprensión es fuerte.

Ahora, es posible transferir el pensamiento del marco de referencia del análisis de varianza al marco de referencia de la regresión múltiple. ¿Es posible obtener $h^2 = .75$ de forma "directa"? La variable independiente *métodos* se considera como la pertenencia en

TABLA 33.1 *Datos ficticios y resultados del análisis de varianza de un factor con tres grupos experimentales*

A_1	A_2	A_3
4	7	1
5	8	2
6	9	3
7	10	4
8	11	5

Fuente	*gl*	*sc*	*cm*	*F*
Entre grupos	2	90.0	45.0	18.0 ($p < .01$)
Dentro de grupos	12	30.0	2.5	
Total	14	120		

TABLA 33.2 *Distribución de los datos y cálculos de regresión (datos de la tabla 33.1)*

	Y	X_1	X_2
	4	1	0
A_1	5	1	0
	6	1	0
	7	1	0
	8	1	0
A_2	7	0	
	8		1
	9	0	1
	10	0	1
	11	0	1
A_3	1	0	
	2	0	0
	3	0	0
	4	0	0
	5	0	0
Σ:	90	5	5
M:	6	.3333	.3333
Σ^2:	660	5	5

los tres grupos experimentales, A_1, A_2 y A_3, que se expresa con los números 1 y 0: si un sujeto es miembro de A_1, se le asigna un 1; si se trata de un miembro de A_2 o de A_3, se le asigna un 0. O se pueden asignar números 1 a los sujetos pertenecientes a A_2 y 0 a los miembros de los otros dos grupos. Los resultados serán básicamente los mismos. De hecho, si se utilizan cualesquiera dos números diferentes, por ejemplo 1 y 10 o 31 y 5, o cualquier serie de dos números aleatorios, los resultados básicos serán los mismos. Sin embargo, la asignación de los números 1 y 0 posee ventajas interpretativas que se mencionarán posteriormente (véase Cohen y Cohen, 1983), y casi siempre funcionan mejor con los programas estadísticos computacionales. La distribución de los datos del análisis de regresión de los resultados de la tabla 33.1 se presenta en la tabla 33.2. Considere como un solo conjunto de puntuaciones las 15 medidas de la variable dependiente en la columna denominada Y. Haga lo mismo con las "puntuaciones" de X_1 y X_2, con excepción de que los números 1 y 0 indican la pertenencia al grupo. A los miembros de A_1 se le asignaron números 1 en la columna X_1; mientras que a los miembros de A_2 y A_3 se les asignó 0 (segunda columna). A los miembros de A_2 se les asignó 1 en la tercera columna, X_2; mientras que a los miembros de A_1 y A_3 se les asginó 0. Se podría plantear la pregunta: ¿en qué parte de la tabla está A_3? Al codificar grupos experimentales, sólo hay $k - 1$ vectores codificados (columnas), donde k es igual al número de tratamientos experimentales (en este caso $k = 3$). Expresado de otra forma, sólo hay un vector codificado (columnas) para cada grado de libertad. Recuerde, de la exposición previa sobre el análisis de varianza, que los grados de libertad entre grupos eran $k - 1$. En tal caso existen tres tratamientos, A_1, A_2 y A_3, y $k = 3$. Por lo tanto son $k - 1 = 2$ vectores codificados. Los vectores de los números 1 y 0 se llaman *variables prototipo o dummy* (véase Suits, 1967, para mayores detalles sobre las variables dummy). Para encontrar una exposición completa sobre las variables codificadas en el análisis de regresión múltiple, consulte los capítulos 6 y 7 de Kerlinger y Pedhazur (1973) o Pedhazur (1966), donde se expresan los tres tratamientos experimentales de forma completa.

TABLA 33.3 *Sumas de cuadrados y productos cruzados (datos de la tabla 33.2)*[a]

	x_1	x_2	y
x_1	3.3333	−1.6667	0
x_2		3.3333	15.0000
y			120.0000

[a] Los valores sobre la diagonal son las sumas de cuadrados de desviación: Σx_1^2, Σx_2^2 y Σy_i^2. Los tres valores restantes que están por arriba de la diagonal son los productos cruzados de desviación: $\Sigma x_1 x_2$, $\Sigma x_1 y$ y $\Sigma x_2 y$.

Ahora realice un análisis de regresión múltiple con los datos de la tabla 33.3, tal como se hizo en el capítulo 32. Las sumas de cuadrados y los productos cruzados necesarios para el análisis se incluyen en la tabla 33.3. Por ejemplo:

$$\Sigma x_1^2 = (1^2 + 1^2 + \ldots + 0^2) - \frac{5^2}{15} = 5 - 1.6667 = 3.3333$$

$$\Sigma x_2 y = (0)(4) + (0)(5) + \ldots + (0)(5) - \frac{(5)(90)}{15} = 45 - 30 = 15$$

$$\Sigma x_1 x_2 = (1)(0) + (1)(0) + \ldots + (0)(0) - \frac{(5)(5)}{15} = 0 - 1.6667 = -1.6667$$

Para calcular la regresión y las sumas de cuadrados residuales se utilizan las fórmulas 32.7 y 32.8 del capítulo 32 (presentadas aquí con la numeración de este capítulo):

$$sc_{reg} = b_1 \Sigma x_1 y - b_2 \Sigma x_2 y \tag{33.1}$$

$$sc_{res} = sc_t - sc_{reg} \tag{33.2}$$

En la tabla 33.3 se incluyen todos los valores anteriores, con excepción de b_1 y b_2, los coeficientes de regresión, y a, la constante de la intersección. Existen varias formas para calcular las b, pero están fuera del alcance de la presente exposición. Por lo tanto, deberán aceptarse por confianza: $b_1 = 3$ y $b_2 = 6$. La constante de la intersección a se calcula de la siguiente manera:

$$a = \bar{Y} - b_1 \bar{X}_1 - b_2 \bar{X}_2 \tag{33.3}$$

$$= 6 - (3)(.3333) - (6)(.3333) = 3$$

Las sumas de los productos cruzados se muestran en la tabla 33.3: $\Sigma x_1 y = 0$ y $\Sigma x_2 y = 15$. Sustituyendo 33.1 en 33.2 se obtiene:

$$sc_{reg} = (3)(0) + (6)(15) = 90$$

$$sc_{res} = 120 - 90 = 30$$

Para calcular R^2 se utiliza la fórmula 32.11 del capítulo 32 (con un nuevo número):

$$R^2 = \frac{sc_{reg}}{sc_t} = \frac{90}{120} = .75$$

$$R = \sqrt{.75} = .8660 \tag{33.4}$$

Por último, se calcula la razón F por medio de la fórmula 32.13, (con un nuevo número):

$$F = \frac{R^2/k}{(1 - R^2)/(N - k - 1)} \tag{33.5}$$

donde k es igual al número de variables independientes y N es igual al número de casos. Sustituyendo:

$$F = \frac{.75/2}{(1 - .75)/(15 - 2 - 1)} = \frac{.375000}{.020833} = 18$$

Se puede tomar prestada otra fórmula de F del capítulo previo:

$$F = \frac{sc_{reg}/gl_1}{sc_{res}/gl_2} = \frac{sc_{reg}/k}{sc_{res}/(N - k - 1)} = \frac{90/2}{30/(15 - 2 - 1)} = \frac{45}{2.5} = 18$$

Después se verifica esta razón F en la tabla respectiva (véase Kerlinger y Pedhazur, 1973, apéndice D o C de este libro), con $gl = 2, 12$. La cifra para $p = .05$ es 3.88, y para $p = .01$ es 6.93. Puesto que la F de 18 calculada antes es mayor que el valor 6.93 de la tabla, la regresión es estadísticamente significativa, y la R^2 también lo es. Note que aun cuando A_3 no fue codificada —no posee vector codificado propio— su media es fácilmente recuperada al sustituir los 0 de X_1 y X_2.

Aun cuando se ha demostrado que el análisis de regresión múltiple logra lo mismo que el análisis de varianza, ¿puede decirse que existe alguna ventaja real en el uso del método de regresión? En realidad, los cálculos son más complicados. Entonces, ¿por qué hacerlo? La respuesta es que con el tipo de datos del ejemplo anterior no existe una ventaja práctica, más allá de la belleza estética y la aclaración conceptual. Pero cuando los problemas de investigación son más complejos —por ejemplo, cuando se involucran interacciones, covariables (como puntuaciones de pruebas de inteligencia), variables nominales (como género, clase social), variables continuas y componentes no lineales (X^2, X^3)— el procedimiento de la regresión tiene ventajas definitivas. De hecho, muchos problemas analíticos de investigación que el análisis de varianza no puede manejar con facilidad, o en lo absoluto, pueden resolverse con bastante facilidad mediante el análisis de regresión múltiple. El análisis factorial de varianza, el análisis de covarianza y, de hecho, todas las formas del análisis de varianza, también se pueden llevar a cabo con el análisis de regresión. Como no es propósito de los autores la enseñanza de la estadística ni de los mecanismos de análisis, se refiere al lector a las explicaciones adecuadas, como las que han sido citadas previamente en este capítulo. Sin embargo, en la siguiente sección se explicará la naturaleza de métodos sumamente importantes de codificación de variables y su utilidad en el análisis.

Codificación y análisis de datos

Antes de extender la explicación sobre la regresión múltiple y el análisis de varianza, es necesario conocer algo sobre las diferentes formas de codificación de los tratamientos experimentales, para el análisis de regresión múltiple. Un código es un conjunto de símbolos asignados a un conjunto de objetos por diversas razones. En el análisis de regresión múltiple, la codificación es la asignación de números a los miembros de una población o muestra,

para indicar la pertenencia a un grupo o subconjunto, de acuerdo con una regla determinada por un medio independiente. Cuando alguna característica o aspecto de los miembros de una población o muestra se define de forma objetiva, entonces es posible crear un conjunto de pares ordenados, cuyo primer miembro constituye la variable dependiente, Y, y el segundo miembro es el indicador numérico de los subconjuntos o de la pertenencia al grupo.

En la explicación anterior sobre la codificación de los tratamientos experimentales en la regresión múltiple análoga al análisis de varianza de un factor, se utilizaron los números 1 y 0. Los vectores de los 1 y de los 0 están correlacionados. Por ejemplo, en la tabla 33.3 la suma de los productos cruzados, $\Sigma x_1 x_2$, es –1.6667, y $r_{12} = -.50$. Dichos códigos 1 y 0 o *dummy* funcionan bastante bien. También es posible utilizar otras formas de codificación. Una de ellas, la codificación de los efectos, consiste en asignar {1, 0, –1} o {1, –1} a los tratamientos experimentales. Aunque se trata de un método útil, sólo se comentará brevemente.

Para aclarar algunos aspectos, la codificación de la tabla 33.2, una regresión múltiple análoga al análisis de varianza de un factor de los datos de la tabla 33.1, con tres grupos experimentales o tratamientos, se presenta en la tabla 33.4. Bajo el encabezado "Dummy" se presenta la codificación dummy de la tabla 33.4, utilizando sólo dos sujetos por grupo experimental. Como existen dos grados de libertad, o $k - 1 = 3 - 1 = 2$, hay dos vectores de columna denominados X_1 y X_2. Ya se explicó la asignación de la codificación dummy: un "1" indica que un sujeto es miembro del grupo experimental junto al que se ha colocado el 1, y un 0 indica que el sujeto no es miembro del grupo experimental.

Bajo la columna de "Efectos", la codificación parece ser {1, 0, –1}. La codificación de los efectos es prácticamente la misma que la codificación dummy —de hecho, se denominó codificación dummy— con excepción de que a un grupo experimental, por lo general al último, siempre se le asignan los números –1. Si las n de los grupos experimentales son iguales, entonces las sumas de las columnas de los códigos son iguales a cero. Sin embargo, los vectores no carecen de correlación de forma sistemática. La correlación entre las dos columnas bajo "Efectos" en la tabla 33.4, por ejemplo, es 50. (Contraste esto con la correlación entre las columnas de los códigos "Dummy": $r = -.50$.)

Cada uno de estos sistemas de codificación tiene sus propias características. Dos de las características de la codificación dummy se comentaron en la sección previa. Por otro lado, una de las características de la codificación de efectos es que la constante de la intersección, a, generada por el análisis de regresión múltiple, igualará a la gran media, o M_1,

TABLA 33.4 *Ejemplos de codificación dummy, de efectos y ortogonal de los tratamientos experimentales*[a]

	Dummy		*Efectos*		*Ortogonal*	
Grupos	X_1	X_2	X_1	X_2	X_1	X_2
A_1	1	0	1	0	0	2
	1	0	1	0	0	2
A_2	0	1	0	1	–1	–1
	0	1	0	1	–1	–1
A_3	0	0	–1	–1	1	–1
	0	0	–1	–1	1	–1
		$r_{12} = -.50$		$r_{12} = .50$		$r_{12} = .00$

[a] En la codificación dummy, A_3 es un grupo control. En la codificación ortogonal, A_2 se compara con A_3, y A_1 se compara con A_2 y A_3, o $(A_2 + A_3)/2$.

de Y. Para los datos de la tabla 33.2 la constante de la intersección es 6.00, que es la media de todas las puntuaciones de Y.

La tercera forma de codificación es la ortogonal; también se denomina codificación de "contrastes", aunque algunas codificaciones de contrastes pueden no ser ortogonales. Como su nombre lo indica, los vectores codificados son ortogonales o no relacionados. Si el principal interés de un investigador son los contrastes específicos entre medias más que en la prueba F en general, la codificación ortogonal puede proporcionar los contrastes requeridos. En cualquier conjunto de datos es posible realizar una cantidad de contrastes. Por supuesto, ello es particularmente útil en el análisis de varianza. La regla es que sólo se hacen los contrastes que son ortogonales entre sí, o independientes. Por ejemplo, en la tabla 33.4 la codificación del último conjunto de vectores es ortogonal: cada uno de los vectores totaliza cero y la suma de sus productos es cero, o

$$(0 \times 2) + (0 \times 2) + (-1)(-1) + \ldots + (1)(-1) = 0$$

r_{12} también es igual a cero.

Suponga que en lugar de la codificación dummy de la tabla 33.2, ahora se utiliza la codificación ortogonal y que también se decide probar A_2 contra A_3, o $M_{A2} - M_{A3}$, y que también se prueba A_1 contra A_2 y A_3, o $M_{A1} - (M_{A2} + M_{A3})/2$. Entonces X_1 se codifica (0, −1, 1) y X_2 se codifica (2, −1, −1), como lo indica la codificación ortogonal de la tabla 33.4. El lector interesado en el análisis de varianza puede seguir dichas posibilidades al consultar a Cohen y Cohen (1983) o a Kerlinger y Pedhazur (1973).

Sin importar qué codificación se utilice, R^2, F, las sumas de cuadrados, los errores estándar de estimación y las Y predichas serán iguales (las medias de los grupos experimentales). La constante de la intersección, los pesos de regresión y las pruebas t de los pesos b serán diferentes. Estrictamente hablando, no es posible recomendar un método sobre otro; cada uno tiene sus propósitos. En un inicio, quizá sea pertinente que el estudiante utilice el método más simple, la codificación dummy o los números 1 y 0. No obstante, poco después se debe usar la codificación de los efectos. Al final, se puede probar y dominar la codificación ortogonal. Antes de utilizar la codificación ortogonal en cualquier grado, el estudiante debe estudiar el tema de la comparación de medias (véase Hays, 1994).

El uso más simple de la codificación constituye la indicación de variables nominales, en particular las de tipo dicotómico. Algunas variables son dicotomías "naturales": género, escuela pública-escuela privada, con convicción-sin convicción, voto a favor-voto en contra. Todas ellas pueden calificarse (1, 0) y los vectores resultantes se analizan como si fueran vectores de puntuación continua. Sin embargo, la mayoría de las variables son continuas, o lo son potencialmente, aun cuando pueden ser siempre tratadas como dicotómicas. En cualquier caso, el uso de vectores (1, 0) para las variables dicotómicas en la regresión múltiple es sumamente útil.

Con variables nominales que no son dicotómicas aun se pueden utilizar los vectores (1, 0). Tan sólo se crea un vector (1, 0) para cada subconjunto, pero sólo uno de una categoría o partición. Suponga que la categoría A se divide en A_1, A_2, A_3, por ejemplo, protestante, católico, judío. Entonces, se crea un vector para protestantes, a cada uno de los cuales se les asigna un 1, a los católicos y judíos se les asigna 0. Se crea otro vector para católicos: a cada católico se le asigna un 1; a los protestantes y judíos se les asigna 0. En efecto sería redundante crear un tercer vector para los judíos. El número de vectores es $k - 1$, donde k es igual al número de subconjuntos de la partición o categoría.

Aunque algunas veces es conveniente o necesaria, la partición de una variable continua en una dicotomía o tricotomía descarta información. Si, por ejemplo, un investigador dicotomiza inteligencia, etnocentrismo, cohesión de grupos o cualquier otra variable que

pueda medirse con una escala que tenga intervalos inclusive aproximadamente iguales, se estaría descartando información potencialmente valiosa. Reducir un conjunto de valores con un rango relativamente amplio a una dicotomía, implica reducir su varianza y, por lo tanto, su posible correlación con otras variables. Por lo tanto, una buena regla del análisis de datos de investigación es: no reducir variables continuas a variables discretas (dicotomías, tricotomías, etcétera), a menos que se esté obligado a hacerlo debido a circunstancias o a la naturaleza de los datos (seriamente sesgados, bimodales, etcétera).

Análisis factorial de varianza, análisis de covarianza y análisis relacionados

Es por medio del análisis factorial de varianza, el análisis de covarianza y las variables nominales que se empiezan a apreciar las ventajas del análisis de regresión múltiple. Aquí se hará algo más que comentar el uso de los vectores codificados en el análisis factorial de varianza. Explicaciones excepcionalmente completas se encuentran en el extenso trabajo de Pedhazur (1996). Sin embargo, se explicará la razón básica de por qué el análisis de regresión múltiple con frecuencia resulta más conveniente que el análisis factorial de varianza.

La dificultad subyacente en investigación y análisis consiste en que las variables independientes de interés estén correlacionadas. Sin embargo, el análisis de varianza supone que no están correlacionadas. Si, por ejemplo, se tienen dos variables experimentales independientes, y los sujetos se asignan aleatoriamente a las casillas de un diseño factorial, se puede suponer que las dos variables independientes no están correlacionadas, por definición. Además, es apropiado usar un análisis factorial de varianza. Sin embargo, si se tienen dos variables independientes no experimentales y las mismas dos variables experimentales, no es posible asumir que las cuatro variables independientes no estén correlacionadas. A pesar de que existen formas para analizar tales datos con un análisis de varianza, éstos son engorrosos y un tanto forzados. Más aún, si existen n desiguales en los grupos, el análisis de varianza se vuelve todavía más inapropiado, a causa de que n desiguales también introducen correlaciones entre las variables independientes. Por otro lado, el procedimiento analítico de la regresión múltiple toma conocimiento, por así decirlo, de las correlaciones entre las variables independientes, así como entre las variables independientes y la variable dependiente. Esto significa que la regresión múltiple puede analizar —de forma separada o conjunta— tanto los datos experimentales como los no experimentales, de manera efectiva. Además, las variables continuas y categóricas pueden utilizarse de manera conjunta.

Cuando los sujetos se han asignado aleatoriamente a las casillas de un diseño factorial, y lo demás permanece igual, no se deriva demasiado beneficio del uso de la regresión múltiple. Pero cuando las n de las casillas son desiguales, y se desea incluir una, dos o más variables de control —como inteligencia, género y clase social— o cuando el análisis involucra el uso de variables continuas, entonces debe utilizarse la regresión múltiple. Este punto es de gran importancia. En el análisis de varianza, añadir variables de control resulta difícil y absurdo. Sin embargo, en la regresión múltiple la inclusión de tales variables es fácil y natural: cada una es sencillamente otro vector de puntuaciones, ¡otra X!

Análisis de covarianza

El análisis de covarianza (y no el análisis estructural de covarianza, que se estudiará posteriormente) es un ejemplo particularmente bueno del valor de un modelo de regresión

múltiple, ya que es difícil y pesado desde el marco conceptual del análisis de varianza, y fácil y completamente comprendido y logrado en un marco de referencia de regresión. Lo que el análisis de covarianza hace en su aplicación tradicional (véase Hays, 1994) es probar la significancia de las diferencias entre medias, después de tomar en cuenta o controlar diferencias individuales iniciales respecto a una *covariable*, es decir, una variable que está correlacionada con la variable dependiente. (Dicha correlación se toma en cuenta.) No obstante, en el modelo de regresión múltiple la influencia de la covariable está controlada tal y como si hubiera cualquier variable independiente, cuya influencia sobre la variable dependiente tuviera que controlarse. La covariable puede ser un pretest o una variable cuya influencia debe ser "eliminada" estadísticamente.

Estudios a gran escala realizados por Prothro y Grigg (1960) y McClosky (1964) encontraron que el grado de acuerdo de las personas acerca de aspectos sociales se incrementa conforme el aspecto se vuelve más abstracto. Suponga que un científico en política considera que el autoritarismo está muy involucrado en esta relación, que cuanto más autoritaria sea la persona, mayor acuerdo mostrará con afirmaciones sociales abstractas. Para estudiar la relación entre lo abstracto y el grado de acuerdo, el investigador tendrá que controlar el *autoritarismo*. En otras palabras, el científico político está interesado en estudiar la relación entre qué tan abstracto es un aspecto y las afirmaciones, por un lado, y el grado de acuerdo con dichos aspectos y las afirmaciones, por el otro. Hasta aquí el interés no se centra en el autoritarismo ni en el grado de acuerdo. El interés principal es *controlar la influencia del autoritarismo sobre el grado de acuerdo. El autoritarismo es la covariable.*

El científico político diseña tres tratamientos experimentales, A_1, A_2 y A_3, que son materiales con diferentes niveles de abstracción. Se obtienen las respuestas de 15 sujetos que han sido asignados aleatoriamente a los tres grupos experimentales, cinco en cada grupo. Antes de que el experimento empiece, el investigador aplica la escala *F* (de autoritarismo) a los 15 sujetos y utiliza tales medidas como la covariable. El objetivo es controlar la posible influencia del autoritarismo sobre el grado de acuerdo. Se trata de un análisis bastante directo de un problema de covarianza, donde se prueba la significancia de las diferencias entre las tres medias del grado de acuerdo, después de corregir las medias de la influencia del autoritarismo y tomando en cuenta la correlación entre autoritarismo y el grado de acuerdo. Ahora se realiza el análisis de covarianza usando el análisis de regresión múltiple.

En la tabla, los datos se presentan en la forma acostumbrada para el análisis de covarianza. En el análisis de covarianza se realizan análisis de varianza separados con las puntuaciones *X*, con las puntuaciones *Y* y con los productos cruzados de las puntuaciones de *X* y *Y*, *XY*. Después, mediante un análisis de regresión, se calculan las sumas de cuadra-

TABLA 33.5 *Análisis de un problema ficticio de covarianza con tres grupos experimentales y una covariable*

Tratamientos					
A_1		A_2		A_3	
X	*Y*	*X*	*Y*	*X*	*Y*
12	12	6	9	12	15
11	12	9	9	10	12
10	11	11	13	4	9
12	10	14	14	4	8
10	12	2	5	8	11

dos y los cuadrados medios de los errores de estimación del total y dentro de grupos y, finalmente, los ajustados entre grupos. Puesto que el interés aquí no es el procedimiento usual del análisis de covarianza, no se realizan estos cálculos. En su lugar, se procede de inmediato al enfoque de regresión múltiple para el análisis.

En la tabla 33.6 se presentan los datos de la tabla 33.5 ordenados para el análisis de regresión múltiple. Como siempre, hay un vector para la variable dependiente, Y. Un segundo vector, X_1, es la covariable. Los dos vectores restantes, X_2 y X_3, representan los tratamientos experimentales A_1 y A_2. (No es necesario tener un vector para A_3 ya que sólo existe un vector por cada grado de libertad, y únicamente hay dos grados de libertad.)

Un análisis de regresión produce: $R^2_{y.123} = .8612$ y $R^2_{y.1} = .7502$. Para probar la significancia de las diferencias entre las medias de A_1, A_2 y A_3, después de realizar el ajuste para el efecto de X_1, la varianza de Y debida a la covariable se resta de la varianza total explicada por la regresión de Y a partir de las variables X_1, X_2 y X_3: $R^2_{y.123} - R^2_{y.1}$. Después se prueba el producto:

$$F = \frac{(R^2_{y.123} - R^2_{y.1})/(k_1 - k_2)}{(1 - R^2_{y.123})/N - k_1 - 1)} \tag{33.6}$$

donde k_1 es igual al número de variables independientes asociadas con $R^2_{y.123}$, la R^2 mayor, y es igual al número de variables independientes asociadas con $R^2_{y.1}$, la R^2 menor. Así, sustituyendo los valores se obtiene:

$$F = \frac{(.8612 - .7502)/(3 - 1)}{(1 - .8612)/(15 - 3 - 1)} = \frac{.0555}{.0126} = 4.405$$

que, con 2 y 11 grados de libertad, es significativa al nivel .05. (Note que un análisis de varianza ordinario de tres grupos, sin tomar en cuenta la covariable, produce una razón F no significativa.) $R^2_{y.23}$, o la varianza de Y explicada por la regresión a partir de las variables dos y tres (los tratamientos experimentales), después de permitir la correlación de la variable

TABLA 33.6 *Análisis de covarianza de los datos ficticios de la tabla 33.5, ordenados para el análisis de regresión múltiple*[a]

	Y	X_1	X_2	X_3
A_1	12	12	1	0
	12	11	1	0
	11	10	1	0
	10	12	1	0
	12	10	1	0
A_2	9	6	0	1
	9	9	0	1
	13	11	0	1
	14	14	0	1
	5	2	0	1
A_3	15	12	0	0
	12	10	0	0
	9	4	0	0
	8	4	0	0
	11	8	0	0

[a] Y = variable dependiente; X_1 = covariable; X_2 = tratamiento A_1; X_3 = tratamiento A_2.

1 y Y, es de .1110. Aunque no se trata de una relación fuerte, especialmente si se compara con la correlación masiva entre la covariable, el autoritarismo y Y ($r^2_{1y} = .75$), sí tiene consecuencias. Evidentemente, lo abstracto de los aspectos influye en las respuestas de acuerdo: cuanto más abstracto es el asunto, habrá mayor acuerdo. Es poco probable que el autoritarismo tenga una correlación con Y de .87. El ejemplo fue alterado deliberadamente para demostrar cómo una fuerte influencia, como la covariable X, puede controlarse y la influencia de las variables restantes (en este caso los tratamientos experimentales) puede evaluarse. Note que la fórmula 33.6 puede utilizarse en cualquier análisis de regresión múltiple; no está limitada al análisis de covarianza o a otros métodos experimentales.

Entonces, el análisis de covarianza se considera simplemente como una variante del tema del análisis de regresión múltiple, en cuyo caso es más fácil de conceptualizar que el procedimiento más bien elaborado del análisis de varianza —en especial si existe más de una covariable (véase Bruning y Kintz, 1987 o Li, 1957)—. La covariable o no es más que una variable independiente. Además, una variable considerada como covariable en un estudio puede ser considerada fácilmente como una variable independiente en otro estudio.

Análisis discriminante, correlación canónica, análisis multivariado de varianza y análisis de ruta

La correlación canónica y el análisis discriminante se dirigen a dos importantes preguntas de investigación: ¿Cuál es la relación entre dos conjuntos de datos con independientes y dependientes? ¿Cómo asignar mejor a los individuos a los grupos, con base en numerosas variables? El análisis de correlación canónica se refiere a la primera pregunta y el análisis discriminante a la segunda. Como se esperaría por el nombre, el análisis multivariado de varianza es la contraparte multivariada del análisis de varianza: se evalúa la influencia de k variables independientes experimentales sobre m variables dependientes. El análisis de ruta constituye más un apoyo gráfico y heurístico que un método multivariado. Como tal, es muy útil, especialmente como ayuda para aclarar y conceptualizar problemas multivariados.

Análisis discriminante

Una función *discriminante* es similar a una ecuación de regresión con una variable dependiente categórica. Sin embargo, cada una posee un propósito diferente. Esta variable dependiente por lo común se presenta en forma de pertenecia al grupo. No obstante, en la regresión múltiple la combinación lineal del predictor o variables independientes sirve para estimar la variable dependiente, la cual en regresión es una medida continua. La mayoría de los investigadores utilizan la regresión múltiple para estimar los valores de la variable dependiente con propósitos de selección. Es decir, si un valor predicho para un conjunto de valores de la variable independiente excede cierto límite, entonces se toma una decisión. El análisis discriminante está involucrado con la clasificación y no necesariamente con la selección. Dado un perfil de puntuaciones en la variable independiente, el análisis discriminante ayuda al investigador a determinar a qué grupo pertenece ese individuo. Algunos investigadores naturalistas han aplicado el método para ayudar a clasificar hallazgos antropológicos de huesos o animales. De la misma manera que en la regresión múltiple, se asume que las variables independientes son continuas, pero que la variable dependiente es categórica. En las situaciones más elementales, la variable dependiente discreta o categórica tiene sólo dos categorías. El problema que la función

discriminante intenta resolver es encontrar un conjunto de coeficientes o pesos, U_i, para las variables independientes (también llamadas variables discriminantes). Se buscan tales pesos para probar si una combinación particular de las variables independientes se asemeja a los miembros de la *categoría 1* o si se asemeja más a los de la *categoría 2*. El objetivo principal consiste en pesar y combinar linealmente las variables independientes, de tal manera que las categorías estén forzadas a ser estadísticamente lo más diferentes que sea posible.

El análisis discriminante responde dos preguntas principales: primero, indica si el conjunto de variables independientes sirve o no para distinguir entre los dos grupos o categorías. La segunda pregunta sólo es importante si la respuesta a la primera pregunta es "sí". La segunda se refiere a la clasificación; indica a qué grupo o categoría debe pertenecer un solo individuo. En otras palabras, la función discriminante separa a los miembros del grupo al máximo. Indica a qué grupo probablemente pertenece cada miembro. Además, también puede hacer una prueba para determinar cuál de las variables independientes explica la diferencia entre los grupos. En síntesis, si se tienen dos o más variables independientes y a los miembros de, por ejemplo, dos grupos, la función discriminante ofrece la "mejor" predicción, en el sentido de los mínimos cuadrados, de la pertenencia "correcta" a un grupo de cada miembro de la muestra.

Algunos investigadores han establecido que el análisis discriminante de dos grupos es igual a la regresión múltiple, con excepción de que la variable dependiente, Y, es dicotómica en lugar de continua. Otros han llegado a aseverar que puede utilizarse cualquier codificación binaria de la variable dependiente (codificación dummy). Sin embargo, ello no es exactamente verdadero. Lindeman, Merenda y Gold (1980) demostraron que los pesos de regresión de la regresión múltiple, b_i, son proporcionales a los pesos de la función discriminante, u_i, si la variable dependiente se codifica como $n_2/(n_1 + n_2)$ para los miembros del grupo 1 y $-n_1/(n_1 + n_2)$ para los miembros del grupo 2.

El análisis discriminante lineal, tal y como lo formuló Fisher (1936), constituye un método apropiado cuando la variable dependiente es categórica. Las variables independientes o predictoras deben medirse en una escala de intervalo. Para probar si existe una diferencia estadísticamente significativa entre grupos, las variables independientes deben distribuirse de manera normal, con varianzas y covarianza iguales. Para utilizar el análisis discriminante de forma adecuada con fines de clasificación, se formulan otras suposiciones acerca de los datos. Un supuesto es que se debe considerar que el perfil de cada individuo tiene la misma posibilidad de estar en cada grupo o categoría. También se debe suponer que el costo de una mala clasificación para cada individuo es el mismo. Estos supuestos, necesarios para la función discriminante, no siempre se cumplen. Como resultado de esto, en años recientes muchos investigadores se alejaron del análisis discriminante a favor de la regresión logística.

Correlación canónica

No es un paso conceptual demasiado grande aquel que va desde el análisis de regresión múltiple con una variable dependiente, hasta el análisis de regresión múltiple con más de una variable dependiente. No obstante, a nivel de cálculos sí es un paso considerable; por lo tanto, no se proporcionarán los cálculos. El análisis de regresión de datos con k variables independientes y m variables dependientes se llama *análisis de correlación canónica*. Se trata de un método que fue desarrollado por Hotelling (1935, 1936). La idea principal es que se forman dos compuestos lineales, por medio de un análisis de mínimos cuadrados: uno para las variables independientes X_i, y otro para las variables dependientes Y_j. La correlación entre estos dos compuestos es la correlación canónica; y, como R, será la máxima correla-

ción posible, dados los conjuntos particulares de datos, debe quedar claro que lo que hasta ahora se ha denominado análisis de regresión múltiple es un caso especial del análisis canónico. En vista de las limitaciones prácticas que tiene el análisis canónico, sería mejor afirmar que el análisis canónico es una generalización del análisis de regresión múltiple.

La correlación canónica puede tener uno o más de los siguientes objetivos:

1. Probar si dos conjuntos de variables están correlacionados o no.
2. Hallar dos conjuntos de pesos o coeficientes, de tal manera que la correlación entre los dos conjuntos esté en su máximo.
3. Buscar en cada conjunto las variables que hagan la mayor contribución a la correlación entre los conjuntos.
4. Predecir los valores en un conjunto de variables, usando valores incluidos en el otro conjunto.

De éstos, quizás el tercer punto sea el más interesante y útil. Por ejemplo, se podría desear determinar cuáles variables basadas en el rendimiento tendrían la mayor relación con un conjunto de medidas de desempeño. La correlación canónica es capaz de brindar dicha información. Después de todo, si se tiene un conjunto de variables basadas en el rendimiento, tales como una batería de pruebas de rendimiento, no todas las pruebas son iguales. Además, no se puede esperar, de manera razonable, que todas realicen la misma contribución. Por lo tanto, resulta lógico considerar a la correlación canónica como un método de correlación de pasos sucesivos que selecciona dos variables, una de cada conjunto de variables, que tengan la relación más fuerte sobre cualquier otro par, después de lo cual, continuará buscando el siguiente mejor par.

En lo que se refiere a los supuestos que deben formularse al aplicar la correlación canónica a los datos, éstos no son tan importantes si no se hacen inferencias acerca del estadístico canónico. Si se utiliza tan sólo con propósitos descriptivos, no se debe asumir que los datos provienen de una distribución multinormal o que provienen de una población con varianzas y covarianza comunes. Sin embargo, si se van a hacer inferencias, como en una prueba de significancia estadística, entonces deben cumplirse tales supuestos. Además, tanto las variables independientes como las dependientes deben medirse con una escala de intervalo, o un conjunto se mide con escala de intervalo y el otro en una escala dicótoma.

De forma similar que en la regresión múltiple y en el análisis discriminante, el objetivo de la correlación canónica aquí es encontrar pesos o coeficientes. La diferencia radica en que hay dos conjuntos en lugar de uno: un conjunto para las variables independientes (también llamadas *predictoras*) y otro conjunto para las variables dependientes (también llamadas *criterios*). Los pesos para ambos conjuntos de variables se encuentran para maximizar la correlación entre los dos conjuntos. Así que, a diferencia de la regresión múltiple y el análisis discriminante, la correlación canónica es capaz de producir más de un conjunto de pesos para las variables independientes y dependientes. Sin embargo, el primer conjunto de pesos sería el que explica la mayor cantidad de varianza. Cada combinación lineal de variables (hay una para cada conjunto de variables) con frecuencia se llama *variante canónico* (véase Lindeman, Merenda y Gold, 1980).

Ejemplo de investigación

Bedini, Williams y Thompson (1995) utilizaron la correlación canónica para estudiar la relación entre el desgaste en el empleo y el estrés del rol terapéutico. Dicho estudio trató únicamente con especialistas en recreación terapéutica. Las medidas de desgaste: *agotamiento emocional, despersonalización y realización personal*, se utilizaron como variables dependientes; mientras que las medidas de estrés del rol: ambigüedad del rol y conflicto del rol se

consideraron como variables independientes. Los investigadores encontraron una función que explicó la relación entre los dos conjuntos de variables. Esta función explicó casi el 36 por ciento de la varianza explicada entre los dos conjuntos. Análisis adicionales determinaron que aproximadamente el 53 por ciento de la varianza de la extenuación se explica por las variables del estrés del rol. Los resultados sugieren que la gente que experimenta estrés del rol tiene mayor probabilidad de sufrir desgaste: experimentar agotamiento emocional, despersonalización y un sentido de baja realización personal.

Análisis multivariado de varianza

Como puede sospecharse, el análisis de varianza posee su contraparte multivariada, el *análisis multivariado de varianza*, el cual permite a los investigadores evaluar los efectos de *k* variables independientes sobre *m* variables dependientes. Al igual que su compañero univariado, que ya se examinó con cierto detalle anteriormente, debe utilizarse con datos experimentales. El análisis multivariado de varianza, o MANOVA, es un método íntimamente relacionado con el análisis discriminante con grupos múltiples. La similitud se presenta sólo en su estructura y no necesariamente respecto a dónde debe utilizarse y cuáles sean los supuestos. Tal como la versión univariada del análisis de varianza presentada en un capítulo previo, los diseños utilizados de forma univariada (una variable dependiente), pueden utilizarse con múltiples variables dependientes. En otras palabras, cada participante del estudio se mide más de una vez, de tal manera que la persona tiene, por lo menos, dos medidas dependientes. En algunos casos, pueden ser dos o más variables dependientes diferentes. En otros casos, puede ser la misma variable medida en diferentes momentos, lo cual a menudo se denomina *análisis de varianza de medidas repetidas*. Algunos investigadores han llamado y analizado inapropiadamente los datos de su estudio como un ANOVA de medidas repetidas, cuando debían haberlo realizado utilizando un MANOVA. La razón de ello es que se debe cumplir el requisito de que el componente de error de las puntuaciones sea independiente. Éste es el supuesto de homogeneidad de la varianza, que en muchas ocasiones es difícil de cumplir, especialmente si las variables dependientes no son verdaderas medidas repetidas. Existen pruebas estadísticas disponibles para demostrar dicho supuesto (véase Kirk, 1995).

Sin embargo, el MANOVA posee unos cuantos supuestos propios que podrían cuestionarse. Las varianzas dentro de grupos, medidas para las variables dependientes para cada uno de los grupos en el análisis, deben ser iguales. También, sería necesario suponer que las variables dependientes se distribuyen de forma multivariada normal. Las pruebas sobre la normalidad multivariada no están lo suficientemente avanzadas. La determinación casi siempre se realiza con series de pruebas poco sistemáticas.

El análisis multivariado de varianza funciona mejor cuando se cumplen los supuestos y también cuando existe una alta correlación entre las variables dependientes. Si la correlación entre las variables dependientes es baja o cercana a cero, el investigador no logrará nada utilizando el MANOVA. En este caso, es posible calcular ANOVAS separados con cada variable dependiente utilizada como una sola medida de resultado. Si éste fuera el caso, el investigador necesitará ajustar el nivel del error tipo I para compensar posibles errores de este tipo. En el otro extremo, si la correlación entre las variables dependientes es 1.00 o cercana a este valor, entonces se sabe que las dos están midiendo esencialmente lo mismo y son redundantes. Siendo así, sólo se necesita calcular un ANOVA para una de esas variables dependientes.

Se evitarán mayores explicaciones aquí, sólo se dirá que, como en todos o en la mayoría de los análisis multivariados, los resultados del análisis multivariado de varianza algunas veces son difíciles de interpretar. Esto se debe a las dificultades mencionadas antes

para evaluar la importancia relativa de las variables en esta influencia sobre una variable dependiente que, como en el análisis de regresión múltiple, con frecuencia están compuestos en el análisis multivariado de varianza, en la correlación canónica y en el análisis discriminante. Si un efecto de interacción resulta estadísticamente significativo, el proceso para determinar qué variables están involucradas en el efecto de interacción puede ser muy laborioso. Bray y Maxwell (1982) y Pedhazur (1996) ofrecen muy buenas explicaciones sobre el análisis multivariado de varianza. Note también que si existen covariables involucradas, entonces se trata de un MANCOVA.

El estudio de Nemeroff (1995) sobre enfermedad y percepción del contagio utilizó un diseño donde los datos recopilados pueden analizarse por medio de un MANOVA. Tal estudio se llevó a cabo para examinar la forma en que la gente reacciona ante los individuos que padecen una enfermedad contagiosa. A los participantes se les dieron crayolas y cuatro hojas en blanco. Se les pidió que dibujaran el germen de la gripe para diferentes personas específicas: para sí mismos, un amigo, un extraño y una persona que les disgustara. Los dibujos fueron calificados en diversas dimensiones por jueces entrenados. Dichas puntuaciones de evaluación sirvieron como variables dependientes. Las dimensiones *activo, grande* y *complejo* se combinaron en una sola medida: *intensidad*. La dimensión *activo* se calificó con base en qué tan activo o pasivo parecía el germen. La *complejidad* se refería a la cantidad de detalle que el participante ponía a los dibujos. Las tres variables individuales restantes eran *abstracción, alcance* y *felicidad*. La *abstracción* se refería a qué tan personificada era la apariencia del dibujo. El *alcance* se refería a qué tan contenido aparecía el germen, y la *felicidad* medía la percepción de los jueces respecto a qué tan agradable o feliz era la apariencia del germen. La variable independiente en tal estudio era la fuente de los individuos, por ejemplo, el novio, el propio sujeto, un extraño, etcétera.

Por medio del uso del MANOVA, Numeroff encontró que la gente percibe el germen de la gripe de forma diferente cuando proviene de una fuente distinta de contagio. Por ejemplo, los gérmenes propios difirieron de los gérmenes de los extraños en su intensidad. Los gérmenes del novio(a) se percibían como menos enojados, por el color, que los gérmenes de personas no agradables, los cuales resultaron más amenazantes. Se encontró que el germen menos amenazante era el del novio(a).

Análisis de ruta

El desarrollo del análisis de ruta se acredita a Wright (1921). El objetivo era desarrollar un modelo causal para la genética y la biología utilizando correlaciones. Como ya se explicó antes, las correlaciones no implican causalidad. No obstante, Wright fue capaz de utilizarlas de ese modo, ya que sus estudios fueron realizados bajo controles muy estrictos. Wright distinguió entre efectos directos e indirectos utilizando correlaciones y regresión. Una variable X puede tener un efecto directo sobre la variable Y, pero un efecto indirecto sobre la variable Z. Tal efecto se establece al examinar los pesos estandarizados de la regresión (correlaciones) entre X y Y, X y Z y Y y Z. Si las correlaciones entre 'X y Y' y 'Y y Z' son altas, pero la correlación entre X y Z es mínima, entonces se tiene un efecto directo entre X y Y y entre Y y Z, y uno indirecto entre X y Z. La contribución de Wright también incluyó una forma específica del uso de las reglas de trazo en los diagramas de ruta para realizar los cálculos necesarios.

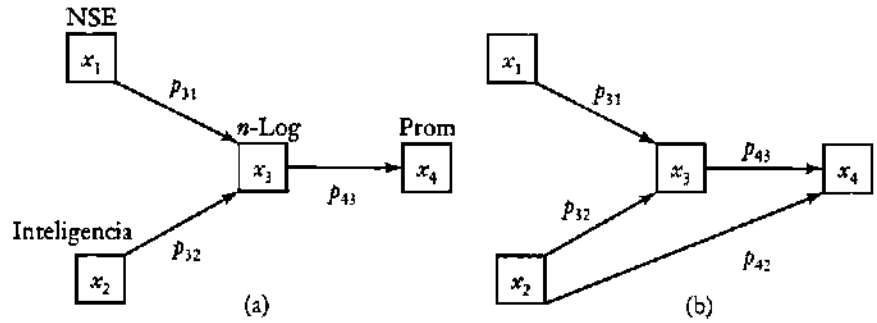
Un redescubrimiento del trabajo de Wright ocurrió en sociología entre la mitad de los años sesenta y el inicio de los setenta. Entonces dichos métodos se volvieron populares con los investigadores en psicología y educación en los años setenta y ochenta. Bentler (1986) ofrece una buena perspectiva histórica de esta transición. Sin embargo, los datos de las ciencias del comportamiento y sociales no son muy similares a los datos que Wright

recolectó y utilizó. Por ende, llamarlo un modelo "causal" es engañoso. Como ya se mencionó, Wright tenía controles muy estrictos para las variables genéticas y de crianza; pero el nivel de control en los estudios en ciencias sociales y del comportamiento resulta mucho menor. Blalock (1972) menciona los requisitos para realizar un análisis de ruta, donde los resultados sean útiles. En la actualidad el análisis de ruta aún funciona como una herramienta de investigación útil para el desarrollo de un modelo conceptual que pueda ser probado de manera empírica. Aunque el término "modelo causal" persiste, en realidad no es causal. Esto será así también cuando se considere el último capítulo sobre el modelamiento de la ecuación estructural. Un libro que ofrece una buena cobertura sobre el análisis de ruta es el de Loehlin (1998).

El análisis de ruta es una forma del análisis de regresión múltiple aplicado que utiliza diagramas de ruta para guiar la conceptualización del problema o para probar hipótesis complejas. Con su uso se calculan las influencias directas e indirectas de las variables independientes sobre la variable dependiente. Dichas influencias se reflejan en los llamados coeficientes de ruta, que en realidad son coeficientes de regresión (beta, b o b). Más aún, es posible probar la congruencia de diferentes modelos de ruta, con los datos observados (véase Pedhazur, 1996). A pesar de que el análisis de ruta ha sido y es un importante método analítico y heurístico, es dudoso que continúe utilizándose como ayuda para probar la congruencia que tienen los modelos con los datos obtenidos. Más bien, su valor será el de un método heurístico que ayude a la conceptualización y a la formación de hipótesis complejas. Sin embargo, la comprobación de dichas hipótesis quizá se realizará con herramientas analíticas más poderosas y más apropiadas para tales comprobaciones. El método que se cubre en el capítulo 35 es, en la actualidad, el mejor método para utilizarse en el análisis y comprobación de hipótesis que surjan a partir de modelos de análisis de ruta. Ahora se estudiará un ejemplo para dar una idea general sobre el método.

Considere los dos modelos, a y b, de la figura 33.1. Suponga que se está tratando de "explicar" el rendimiento o promedio, x_4, en la figura, o Prom. Una persona considera que el modelo a es "correcto"; sin embargo, una segunda persona considera que el modelo b es "correcto". En efecto, el modelo a dice que tanto el nivel socioeconómico NSE, como la inteligencia influyen en x_3, que representa la necesidad de logro (n-Log), y que x_3 influye en x_4, Prom o rendimiento. ¡Qué bien! En otras palabras, la primera persona considera que el modelo a expresa mejor las relaciones entre las cuatro variables. Por otro lado, la segunda persona considera que el modelo b es una mejor representación. Éste añade una influencia directa de x_2, inteligencia, sobre x_4, rendimiento (note las rutas de x_2 a x_4 y de x_2

FIGURA 33.1 x_1 = nivel socioeconómico (NSE); x_2 = inteligencia; x_3 = n-Log o necesidad de logro; x_4 = Prom o promedio de calificaciones (rendimiento)



a x_3 y a x_4). ¿Qué modelo es "correcto"? En el análisis de ruta se prueban los dos modelos utilizando el método del capítulo 35.

Regresión de cresta, regresión logística y análisis logarítmico lineal

Regresión de cresta

La regresión de cresta constituye un método reconocido por los estadísticos aplicados y por los investigadores de las ciencias de ingeniería. Sin embargo, no ha sido popular en la investigación de la ciencia psicológica o del comportamiento. Los creadores del método, Arthur Hoerl y Robert Kennard, publicaron su artículo monumental en 1970. A pesar de la actitud de la psicología hacia su método, el impacto del artículo de Hoerl y Kennard ha sido tan grande que el Instituto para la Información Científica (Institute for Scientific Information) lo denominó como una "cita clásica" (Hoerl, 1995).

La psicología casi siempre asocia el artículo de Price (1977) como la introducción de la psicología a la regresión de cresta. No obstante, el manuscrito bien escrito e informativo de Simon (1975) sobre el tema precedió al de Price por dos años. Además, Bolding y Houston (1974) desarrollaron un programa computacional para realizar la regresión de cresta. Simon (1975) demostró cómo podía usarse el método en estudios de factor humano, donde no se cumplían todos los requisitos de un experimento verdadero. Dichos estudios incluían una o más variables predictoras que estaban *altamente* correlacionadas. Las variables correlacionadas se debían al fracaso para completar un experimento verdadero o a la falta de habilidad del investigador para seleccionar o controlar condiciones experimentales relevantes. Simon llama a este tipo de estudios "no diseñados". Keith (1988) y Price (1977) los denominan "investigación no experimental". Hoerl y Kennard (1970) los consideran "estudios no ortogonales".

La desaprobación de la regresión de cresta por parte de la investigación psicológica quizá se deba en parte a las críticas de Rosenboom (1979) acerca del método. La mayoría de los métodos bayesianos o las técnicas de estimación del sesgo requieren de la intervención y juicio humanos, en lugar de un método estrictamente matemático y analítico, tal como el de los mínimos cuadrados. La regresión de cresta es uno de dichos métodos, y como tal se ha considerado deshonesto. Aun los autores de reconocidos libros sobre estadística, como Draper y Smith (1981) afirman que el método fue muy polémico en los años setenta. Sin embargo, como se estableció en un artículo de Frank y Friedman (1993), la regresión de cresta constituye claramente el mejor método de análisis de regresión en muchas condiciones no experimentales. Diversos autores coinciden con Keith (1988) en que la regresión múltiple es el método de elección cuando se trata de estudios no experimentales. No obstante, la regresión múltiple se debilita con rapidez cuando las variables predictoras están altamente correlacionadas o son colineales. Esto se debe al hecho de que la regresión múltiple, tal como se maneja actualmente en la mayoría de los programas estadísticos, utiliza el método de los mínimos cuadrados. El lector debe notar que el segundo autor de este libro está tomando una posición más bien extrema al respecto. En estudios de investigación donde las variables predictoras están ligera o moderadamente correlacionadas (alrededor de .50 o menos), el uso de un método como la regresión de cresta podría no ser necesario. De hecho, Keith (1999) ha señalado que la regresión múltiple produce estimados bastante estables de los coeficientes de regresión cuando el nivel de colinealidad es moderado.

El problema con los mínimos cuadrados ordinarios (MCO)

Uno de los propósitos del análisis de regresión múltiple es obtener un conjunto de coeficientes o pesos no sesgados, que tengan una mínima cantidad de error de la variable, así como un ajuste razonable con un conjunto existente de datos. Un método popular para hacerlo es el de los mínimos cuadrados ordinarios (MCO), el cual se trata en todo libro de texto de estadística elemental. Aquí lo abordamos en un capítulo anterior. Este método es directo, determinado matemáticamente y no requiere del juicio humano. Además implica la estimación de los coeficientes de regresión, con la limitante de que la suma de cuadrados de la diferencia entre la medida resultante predicha y observada sea la mínima:

$$\Sigma(Y'_i - Y_i)^2 = \text{mínimo para la ecuación } Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \ldots + \beta_n x_n + e$$

En esta ecuación las X son las variables predictoras y la Y es la variable dependiente o criterio. Cuando las variables predictoras son matemáticamente independientes (por ejemplo, las correlaciones entre las X son iguales a *cero*), los coeficientes de regresión estimados son representaciones razonables de los verdaderos coeficientes de regresión, dentro de los límites de las fluctuaciones de muestreo. Cuando las variables predictoras están altamente correlacionadas, los coeficientes individuales de regresión calculados con el método de MCO con frecuencia resultan insatisfactorios. La matriz de variables predictoras altamente correlacionadas se llama *mal condicionada*. Las cualidades predictivas de dicha ecuación generada por medio de los mínimos cuadrados son razonablemente precisas para los datos utilizados para generar la ecuación. No obstante, la aplicación de la misma ecuación de regresión a un nuevo conjunto de datos da como resultado valores predichos pobres para la medida resultante. Es decir, la validación cruzada de la ecuación de regresión es extremadamente pobre. Además, los efectos relativos de los coeficientes individuales de regresión no pueden evaluarse. En otras palabras, los coeficientes de regresión obtenidos por medio de los mínimos cuadrados en variables predictoras correlacionadas quizá no tengan sentido cuando se evalúan en términos del mundo real. Hoerl y Kennard (1970), retomados por Simon (1975), afirmaron que una o más de las siguientes características pueden presentarse en un ajuste de mínimos cuadrados de variables predictoras correlacionadas:

1. los coeficientes de regresión se vuelven demasiado grandes en su valor absoluto,
2. algunos coeficientes llegan a tener el signo equivocado,
3. los coeficientes son inestables; otro conjunto de datos para las mismas variables producirá valores resultantes diferentes, y
4. los pesos individuales de regresión sobrestiman o subestiman el efecto de una variable en particular.

Dos variables altamente correlacionadas resultarán en coeficientes donde una variable recibe un peso grande y la otra recibe un peso pequeño e insignificante. Una explicación más completa sobre los peligros de los MCO se encuentra en Newman (1976).

Hoerl y Kennard (1970) desarrollaron una alternativa al método de regresión múltiple convencional con variables predictivas correlacionadas. El método se creó para permitir a los investigadores evaluar variables en problemas de ingeniería química, donde sería impráctico abandonar variables o crear variables compuestas. Este método, llamado *regresión de cresta* produce una mejor ecuación de predicción que la que se obtendría utilizando los mínimos cuadrados y es mejor debido a que los coeficientes estimados se acercan más a los coeficientes verdaderos, en promedio. Los signos de los coeficientes son más precisos; los coeficientes son más estables, con una mayor probabilidad de repetirse con un

nuevo conjunto de datos, y la medida estimada resultante puede lograrse con un error de cuadrados medios más pequeño. A través de la regresión de cresta los valores eigen se vuelven menos discrepantes.

En esencia, el análisis de regresión de cresta es idéntico a la regresión de MCO, con excepción de que un número pequeño, k, se ha añadido a la diagonal de la matriz de correlación de las variables predictoras. La añadidura de dicho número k a la diagonal hace a la matriz menos mal condicionada. También tiene el efecto de disminuir el error de los cuadrados medios cuando se comparan con los mínimos cuadrados. El mejor valor k es aquel donde los coeficientes de regresión se estabilizan y donde las sumas de cuadrados residuales son bajas. Para encontrar este valor el investigador debe probar diversos valores de k. Un número de investigadores objetaron dicho método *ad hoc* para seleccionar k. Como Simon (1975) demostró, el valor k puede agregarse directamente a la diagonal de la matriz de correlación y, después, someterse a un programa computacional de regresión, o el investigador puede añadir casos dummy a los datos originales (datos aumentados) y seleccionar la opción de intersección cero del programa computacional. El BMDP (Dixon, 1990) tiene un subprograma en su paquete estadístico para realizar la regresión de cresta. Lee (1980) demuestra cómo se puede efectuar lo anterior con programas de regresión que no tienen una opción de intersección cero. Se han realizado diversos estudios para desarrollar una forma más analítica para determinar k. La explicación de dichos estudios ocuparía demasiado espacio en la presente obra. Se pide al lector que consulte el artículo de Simon o el de Draper y Smith (1981).

El precio que se paga por utilizar la regresión de cresta es que los coeficientes de regresión ya no están sin sesgo y que la suma de cuadrados residual (SCR) ya no es mínima. Sin embargo, los beneficios de una regresión de cresta ejecutada apropiadamente llegan a contrarrestar tales desventajas. No obstante, como todos los métodos estadísticos, la regresión de cresta debe emplearse de forma correcta. Draper y Smith (1981, p. 322) lo exponen de manera más directa con sus palabras:

> La selección de la regresión de cresta no constituye una panacea milagrosa; se trata de una solución de mínimos cuadrados restringida por la suma de alguna información externa acerca de los parámetros. El uso a ciegas de la regresión de cresta sin esta consideración resulta peligroso y engañoso. Si la información externa es sensata y no se contradice con los datos, entonces la regresión de cresta también es sensata.

Ejemplo de investigación

Bee y Beronja (1986) utilizaron la regresión de cresta en un estudio de estudiantes universitarios con un área de estudios principal indeterminada. Los investigadores reunieron los resultados de una prueba de ingreso a la universidad (ACT) junto con el desempeño académico universitario y medidas de personalidad, como motivación y hábitos de trabajo académico. La meta consistía en desarrollar una ecuación de regresión que pudiera predecir el desempeño académico universitario (promedio de calificaciones) utilizando variables de personalidad, variables de experiencia del programa (por ejemplo, nivel de dificultad de los cursos en el área de estudio principal) y las puntuaciones de una prueba de ingreso a la universidad (ACT). Tales variables explicativas o independientes eran colineales. Cuando Bee y Beronja ajustaron los datos por medio del método de regresión ordinaria, encontraron que ninguna de las variables explicativas estaba en relación significativamente con el desempeño académico. No obstante, los estimados de la regresión de cresta proporcionaron resultados muy diferentes. La regresión de cresta encontró que las variables ACT de matemáticas, los hábitos de trabajo académico, la motivación para el éxito y la dificultad de los cursos de matemáticas estaban relacionadas de manera significativa con el desempeño

académico. Bee y Beronja encontraron que $k = .4$ producía los mejores resultados en la regresión de cresta. Ninguno de los pesos de regresión encontrados madiante los mínimos cuadrados ordinarios fue estadísticamente significativo ($p > .05$). Sin embargo, cuatro de los pesos de regresión que fueron determinados por medio de la regresión de cresta resultaron significativos.

Regresión logística

Ya se explicó anteriormente en el libro la regresión múltiple y el análisis de función discriminante. Por lo general, si se tiene una variable dependiente categórica, la recomendación era realizar un análisis de función discriminante, el cual, no obstante, sólo es efectivo si las variables cumplen ciertos supuestos. En algunos estudios de las ciencias sociales y del comportamiento, las variables independientes o predictoras son categóricas o nominales. Cuando esto sucede, el análisis de función discriminante empieza a perder su eficacia en términos de su buen ajuste con los datos. Si se utiliza la regresión múltiple, la ecuación con mejor ajuste quizá sea inaccesible y puede tratarse de una ecuación que no produzca información útil. Después de todo, la regresión múltiple tradicional supone que los datos se miden en una escala de intervalo (o algo cercano a ello) y que sigue una distribución normal.

Un método que va ganando gran popularidad en años recientes es la regresión logística. Su desarrollo parece muy novedoso a los investigadores en el campo de la psicología, la sociología y la educación. La vida y las ciencias médicas lo han utilizado por un periodo mucho más largo. El retraso de las ciencias sociales en el uso de este método es irónico. Durante los años sesenta, los psicólogos y los investigadores sociales del Instituto de Investigación Social (Institute for Social Research, ISR) de la Universidad de Michigan, habían desarrollado los métodos denominados *análisis de clasificación múltiple* (ACM) y el *análisis multivariado de escala nominal* (AMN), ahora conocido como regresión logística. En realidad, la psicología tuvo un temprano encuentro con el método, el cual permaneció latente durante años, con excepción de aquellos afiliados o que conocían el Instituto de Investigación Social.

Andrews, Morgan, Sonquist y Klem (1973) describen una técnica para examinar las relaciones entre variables independientes y una variable dependiente, que es similar a la regresión múltiple. Ellos señalan el problema implicado cuando las variables independientes o predictoras se miden en una escala nominal y ofrecen una solución. El análisis de clasificación múltiple (ACM), como ellos lo llaman, incluye datos tanto nominales como no nominales en las variables predictoras. En la variable dependiente o criterio, los datos se miden con una escala de intervalo o dicotómica. El *análisis multivariado de escala nominal* (Andrews y Messenger, 1973) es una extensión del ACM. Permite el uso de variables dependientes de escala nominal con más de dos categorías. De acuerdo con la nomenclatura actual, el ACM se denomina *regresión logística;* y el AMN se conoce como *regresión logística policotómica* (véase Dixon, 1990). Se utilizarán los términos más populares en la explicación de este método.

La *regresión logística,* por lo tanto, constituye una técnica para ajustar una superficie de regresión con los datos, en la cual la variable dependiente es dicotómica. En psicología educativa es posible clasificar a los estudiantes con un funcionamiento mental *alto* o *bajo*; o, en referencia a la terapia, se puede clasificar a los pacientes como *con mejoría* y *sin mejoría*, *exitosa* o *no exitosa*. En cada estudio que utiliza una variable dependiente dicotómica, la regresión logística constituye un candidato viable como método de análisis. Sin embargo, se podría plantear la pregunta: "¿Cuál método se debe utilizar: el análisis discriminante o la regresión logística?" Existe controversia respecto a la comparación del análisis discrimi-

nante con la regresión logística. Press y Wilson (1978) reportaron aquellas situaciones donde el análisis discriminante funciona bastante bien; es decir, cuando los datos cumplen los supuestos. No obstante, cuando no se cumplen los supuestos, la regresión logística representa una forma superior de análisis. La regresión logística requiere que se cumplan menos supuestos; pero no constituye una panacea para los datos recolectados con diseños de investigación cuestionables. El análisis discriminante produce con facilidad una probabilidad de éxito que cae fuera del rango del 0 al 1, lo cual no es aceptable. Por otro lado, la regresión logística no produce probabilidades más allá de entre 0 y 1. Ambos producen estimados de regresión y ambos son capaces de clasificar individuos. En la regresión logística el investigador obtiene una ganancia adicional: el coeficiente o peso de regresión puede transformarse en *razones de probabilidad*, un estadístico útil que se describió en el capítulo 10. Es útil al ofrecer al investigador ideas respecto a lo que está sucediendo dentro de los datos.

Con una comparación directa entre el análisis discriminante y la regresión logística, esta última funciona mejor cuando las variables no son normales. Además, la regresión logística no se ve tan afectada, como el análisis discriminante, cuando se incluyen variables sin significado dentro del análisis. Lo anterior implica variables dicotómicas o variables que se han sometido a codificación dummy. El uso de variables dicotómicas es muy común en la investigación de las ciencias del comportamiento. Así, si los propios datos de investigación contienen variables categóricas o de escala nominal, o si existe alguna duda razonable respecto a alguna de las variables, quizá sea mejor utilizar la regresión logística en lugar del análisis discriminante.

Podría realizarse una comparación similar entre la regresión logística y la regresión múltiple ordinaria utilizando variables independientes y dependientes dicotómicas. Con la regresión múltiple los valores predichos caerían fuera del rango 0-1. Además, si las varianzas calculadas de la variable dependiente fluctuaran con valores de las variables dependientes, como tener más números 1 que 0 para un nivel e igual número de 0 y 1 en otro, el análisis producirá una varianza grande. Como resultado, se violarán los supuestos de homogeneidad de varianza y de normalidad. Sin embargo, existen situaciones donde la regresión múltiple con una variable dependiente dicotómica daría buenos resultados (véase Cox y Wermuth, 1992).

Un ejemplo de investigación

Los datos de este ejemplo son cortesía de Dorothy Scattone de la Universidad del Sur de Mississippi. Su estudio trató de las percepciones de dos grupos diferentes de asiáticos, hacia las discapacidades físicas y mentales. Un grupo de asiáticos había nacido y crecido en un país asiático; mientras el otro grupo se componía de asiáticos nacidos en Estados Unidos. Los participantes fueron 215 estudiantes universitarios que respondieron un número de preguntas respecto a ciertas discapacidades como el síndrome de Down o su nivel de aceptación hacia personas con asma o cicatrices faciales, etcétera. Las variables se midieron en una escala de calificación de 5 puntos, donde el 5 significaba una alta aceptación y el 1 una baja aceptación.

Los resultados de este análisis indican cuáles variables discriminaron entre los asiáticos nacidos en Estados Unidos y los nacidos en el extranjero; además señalan las respectivas probabilidades. Respecto a la variable *tartamudeo*, se les pidió a los participantes que indicaran su nivel de aceptación de una persona con problemas de habla, específicamente con tartamudeo, donde el 1 implicaba la no aceptación y el 5 una aceptación total. Con la aceptación total, el sujeto afirma que está de acuerdo con tener a la persona como un miembro de la familia a través del matrimonio. Puesto que esta variable fue significativa, indica que los asiáticos nacidos en Estados Unidos y los nacidos en el extranjero difieren

en su respuesta. La razón de probabilidad para las variables es una forma diferente de hablar sobre las probabilidades. La ventaja que la razón de probabilidad tiene sobre la prueba de significancia consiste en que la primera casi no se ve afectada por el ramaño de la muestra. Para la variable *tartamudeo* la razón de probabilidad fue .4516, lo cual indica que los asiáticos nacidos en el extranjero tienen .4516 veces menos probabilidades de aceptar a las personas con tartamudeo, que los asiáticos nacidos en Estados Unidos o, en otras palabras, los asiáticos nacidos en Estados Unidos tienen 2.2 veces más probabilidades de aceptar a una persona con dicha incapacidad del habla, que los asiáticos nacidos en el extranjero (1/.4516).

Además de esta información, el análisis de regresión logística también proporciona una medida sobre qué tan precisa es la ecuación de regresión, en términos de clasificación. Con tales datos, la ecuación de regresión logística fue capaz de clasificar correctamente al 76.74 por ciento de los casos. La ecuación resultó, de forma general, más precisa al predecir a los asiáticos nacidos en el extranjero (92.02 por ciento) que a los asiáticos nacidos en Estados Unidos (28.85 por ciento). Existen otros estadísticos, como la prueba Wald, que se asocian con un resultado de regresión logística; aunque no se estudiarán aquí; en su lugar, se referirá al lector a libros como el de Hosmer y Lemeshow (1989) o el de Shoukri y Edge (1996).

Tablas de contingencia de múltiples factores y análisis log-lineal

Es adecuado presentar esta sección después de la regresión logística a causa de que el tema que se inicia a continuación trata de manera exclusiva con datos categóricos. En la regresión logística se tiene una variable dependiente dicotómica y variables independientes categóricas y de intervalo. En las tablas de contingencia de múltiples factores se trata sólo con datos categóricos. Los análisis de las tablas de contingencia de múltiples factores son importantes porque muchos de los datos utilizados por los científicos sociales y del comportamiento son categóricos. El empleo de los métodos tradicionales de análisis de varianza y regresión múltiple para analizar datos categóricos no funciona bien en muchos casos.

En el capítulo 10 se estudiaron las tablas de contingencia unidimensional y bidimensional. En aquel momento, se introdujo de forma breve el concepto de las tablas de múltiples factores. Tradicionalmente muchos investigadores estudiaban las tablas de contingencia de múltiples factores observando una serie de tablas de dos factores. Los cálculos son relativamente directos y el investigador puede, por lo general, llegar a alguna conclusión razonable respecto de los datos. También, tienen menor probabilidad de tener casillas con pocos casos o vacías. Uno de los eventos más probables en tablas grandes es la presencia de casillas con pocos casos o vacías, lo cual puede afectar los resultados del análisis. Como consecuencia, muchos investigadores reducen el número de categorías para eliminar este problema. Sin embargo, las series de tablas de contingencia de dos factores, utilizadas para analizar tablas de contingencia de factores múltiples, no permiten captar la existencia de efectos de interacción de orden superior entre las dimensiones. Glick, DeMorest y Hotze (1988) ofrecen un buen ejemplo de cómo analizar una tabla de contingencia de tres factores de forma correcta con una serie de análisis de dos factores. También fueron capaces, a través de una progresión de afirmaciones y análisis lógicos, de llegar a una conclusión acerca del término de interacción de tres factores. Posteriormente, en esta sección, se hará un nuevo análisis de sus datos a la luz de las tablas de contingencia de factores múltiples o del análisis logarítmico lineal. Además, las asociaciones entre las variables difieren más a través del análisis de dos factores que a través de factores múltiples, ya que este último

toma en consideración las otras variables implicadas. Además, el uso de tablas de sólo dos factores no permite la comparación simultánea de todas las asociaciones de pares.

Recuerde que en el análisis de datos categóricos del capítulo 10 se estableció la diferencia entre los valores observados y los valores esperados. Ambos representan frecuencias dentro de cada casilla de una tabla de contingencia. Si las frecuencias esperadas coinciden con las frecuencias observadas, entonces se diría que no hubo relación entre las dos variables categóricas, lo cual sucede así porque las frecuencias esperadas se calculan bajo condiciones de lo que se podría esperar si no hubiera relación entre las dos variables. Por lo tanto, si las frecuencias observadas coinciden con las frecuencias esperadas, es posible concluir que los datos reunidos no encontraron relación alguna. De forma subsecuente, si no hubo coincidencia, entonces se afirma que las dos variables están relacionadas. El análisis de las tablas de contingencia de factores múltiples opera de una forma muy similar. El investigador especifica un modelo que incluya a las variables, como sin interacciones entre tres factores o sólo interacciones específicas entre dos factores. Una vez especificado el modelo, se generan las frecuencias esperadas. Si las frecuencias observadas coinciden con las frecuencias esperadas, entonces se sabe que el modelo elegido se ajusta a los datos observados y los elementos del modelo explican los valores observados. En las tablas de factores múltiples, una de las metas consiste en encontrar las variables que se relacionan con otras variables. Hallar los valores esperados para probar si los valores observados coinciden es más demandante a nivel de los cálculos que en las tablas simples de dos factores. A pesar de que no se realizarán los cálculos, se puede dirigir al lector hacia referencias útiles que demuestran con claridad la operación. Uno de los algoritmos más populares para encontrar los valores esperados fue desarrollado por Deming y Stephan (1940). Una descripción de dicho método puede encontrarse en su artículo original. También se encuentran en la reimpresión de Dover de 1964 del libro de Deming de 1943 o en Dillon y Goldstein (1984), quienes ofrecen un ejemplo del cálculo claro y fácil de seguir. Algunas veces el método de Deming y Stephan se denomina *ajuste iterativo proporcional*. De acuerdo con su nombre, el ajuste iterativo proporcional requiere de un estimado inicial de las frecuencias esperadas y después, a través de un número de pasos, se ajustan. En la primera iteración se obtienen los estimados que van a utilizarse en la siguiente iteración. Tal como sucede en la regresión logística, las iteraciones cesan una vez que dos iteraciones sucesivas producen estimados muy cercanos entre sí.

Uno de los beneficios producidos por el método log-lineal es el de la parsimonia. Es decir, con el método logístico lineal el investigador especifica un modelo de términos al cual ajustarse, de manera muy similar a lo que se hace en el análisis de varianza o en la regresión múltiple. El investigador intenta obtener el mejor ajuste posible con el menor número de términos. Considere que m_{ij} represente la frecuencia esperada en la casilla (i,j) de una tabla de contingencia de dos factores. Nombre a una de las variables A y a la otra B. A tendría i categorías y B tendría j categorías. El modelo para esta tabla de contingencia se puede expresar como:

$$\log_e(m_{ij}) = \mu + \mu_{A(i)} + \mu_{B(j)} + \mu_{AB(i,j)}$$

Algunas ocasiones la ecuación se escribe sin los subíndices i y j, los cuales se utilizan para indicar el número de categorías en cada variable. Si la ecuación se escribe sin los subíndices, entonces las categorías están implícitas. Para simplificar, se escribirán las ecuaciones sin referencia directa al número de categorías en cada variable. Se podrá notar que tal ecuación se asemeja a aquella utilizada en el análisis de varianza. Sin embargo, entre las diferencias, el lector debe recordar que en la regresión múltiple o en el análisis de varianza la ecuación se desarrolla para predecir o explicar la variación de un individuo a otro. En otras palabras,

el análisis de varianza y la regresión múltiple hacen un estimado de la variable dependiente para cada caso o individuo. En el log-lineal o tablas de contingencia, la predicción se hace respecto a la frecuencia o categoría de la casilla y no respecto al individuo. A pesar de que ciertos escritores (Bakeman y Robinson, 1994; Fienberg, 1980; Howell, 1997; Kennedy, 1992) sobre tablas de contingencia de factores múltiples y sobre análisis log-lineal hacen analogías con la regresión múltiple y el análisis de varianza, ellos enfatizan esta importante diferencia.

En el análisis log-lineal, si se tuvieran tres variables categóricas, se escribiría como

$$\log_e(m_{ijk}) = \mu + \mu_A + \mu_B + \mu_C + \mu_{AB} + \mu_{AC} + \mu_{BC} + \mu_{ABC}$$

Note que en cada modelo existen términos principales y las interacciones. Cuando todas las posibles combinaciones de los términos están explicadas en la ecuación, el modelo se denomina *saturado*. El estadístico de bondad de ajuste es un cero perfecto que indica que el modelo se ajusta con los datos observados; es decir, los valores observados se ajustan con los valores esperados. Sin embargo, un investigador puede ajustar diferentes modelos con este método de análisis. La guía dictada por la teoría o por información previa sirve para eliminar algunos de los términos. Si se encuentra que los datos observados se ajustan con los valores esperados generados por el nuevo modelo, entonces se ha ajustado exitosamente un modelo más parsimonioso. Los modelos más parsimoniosos con frecuencia se denominan *modelos reducidos* o *modelos no saturados*. Lo que un modelo bien ajustado no saturado indica es que no se necesitan todos los términos del modelo saturado para obtener un ajuste adecuado de los valores observados con los valores esperados.

Los términos de interacción del modelo log-lineal son conocidos como *términos de orden superior*. A mayor número de términos en la interacción, más alto será el término. En el modelo de tres factores presentado anteriormente, el término μ_{ABC} es el término de orden superior; mientras que μ es el término de orden inferior. Esta breve especificación es importante para la explicación de la diferencia entre los modelos jerárquicos y los no jerárquicos. Algunos expertos en el análisis log-lineal para tablas de contingencia han establecido que el modelo jerárquico es el más útil y que los resultados de los modelos no jerárquicos son cuestionables (Bakeman y Robinson, 1994; Howell, 1997). La siguiente explicación se restringirá solamente a los modelos jerárquicos, en los cuales se observan los términos de orden superior como un compuesto de términos de orden inferior. Para calcular μ_{AB} se necesitaría calcular también μ, μ_A y μ_B, los cuales son todos los términos de orden inferior. Así, en los modelos jerárquicos los términos de orden superior se incluyen sólo si los términos de orden más bajo también se incluyen en el modelo. Los modelos no jerárquicos no tienen esta restricción y, como tales, generan resultados difíciles de interpretar.

Puede haber un gran número de modelos no saturados. Conforme se incrementa el número de variables categóricas, también lo hace el número de modelos. En algunos casos se vuelve muy difícil probar todos los modelos. De hecho, el investigador no debe intentar probar todos los modelos con la esperanza de hallar uno que se ajuste a los datos. El modelo debe basarse en la teoría o en una combinación de teoría y hallazgos previos. Tome como ejemplo el estudio de Glick, DeMorest y Hotze (1988), quienes encontraron una interacción de tres factores sin utilizar el llamado análisis log-lineal. Dicho término de interacción de tres factores, μ_{ABC}, en términos log-lineales sería el término de orden superior para los datos. Por lo tanto, el modelo se especificaría con el término de tres factores. O, si se deseara extenderse en el análisis e incluir una cuarta variable categórica, definitivamente se incluiría en el modelo un término para la interacción de tres factores. La decisión sobre cuáles términos se incluirán en el modelo para comprobarlo se denomina *especificación*.

La especificación del modelo para un análisis log-lineal en tablas de contingencia utiliza una notación especial. Se utilizan letras mayúsculas del alfabeto, encerradas en corchetes, para representar el efecto de cada variable de forma separada. Por ejemplo, cuando se habla del efecto de A en una tabla de tres factores, se escribiría $[A]$. En un modelo jerárquico, si se establece $[AB]$, se está refiriendo al modelo:

$$\log_e(m_{ij}) = \mu + \mu_A + \mu_B + \mu_C + \mu_{AB}$$

Si se anota $[A][BC]$ se refiere al modelo:

$$\log_e(m_{ij}) = \mu + \mu_A + \mu_B + \mu_C + \mu_{BC}$$

Si se anota $[ABC]$, entonces se estaría hablando del modelo saturado:

$$\log_e(m_{ij}) = \mu + \mu_A + \mu_B + \mu_C + \mu_{AB} + \mu_{AC} + \mu_{BC} + \mu_{ABC}$$

Bakeman y Robinson (1994) emplean un sistema de notación que difiere ligeramente de éste. El programa computacional que acompaña su libro es muy fácil de usar y parece estar tan bien desarrollado como algunos de los programas disponibles comercialmente. Sin embargo, su programa no imprime los corchetes "[ ]".

El estadístico de bondad de ajuste que se revisó en el capítulo 10 tiene un nombre formal que no se mencionó, que es necesario en este capítulo principalmente porque se presentará otro estadístico de bondad de ajuste. El estadístico revisado en el capítulo 10 es la chi cuadrada de Pearson. Su fórmula es:

$$\chi^2 = \sum\left[\frac{(f_o - f_e)^2}{f_e}\right]$$

Otro estadístico que es casi idéntico a la χ^2 de Pearson es la de razón de probabilidad χ^2. Para distinguir entre ambas, la chi cuadrada de la razón de probabilidad por lo común se anota como G^2. Como lo explica Wickens (1989), las dos son casi idénticas respecto a su aproximación a una distribución de chi cuadrada. La decisión sobre cuál debe usarse es una cuestión de preferencia. Wickens sí menciona que la χ^2 de Pearson es más familiar y más clara a nivel intuitivo, que la razón de probabilidad. Algunos programas computacionales calculan ambas. La ventaja que la chi cuadrada de la razón de probabilidad (G^2) tiene sobre la chi cuadrada de Pearson es a nivel del cálculo. La fórmula de la razón de probabilidad no utiliza las frecuencias esperadas de forma directa. La chi cuadrada de la razón de probabilidad para una tabla de dos factores se presenta como:

$$G^2 = 2\left(\sum f_{0ij}\log f_{0ij} - \sum R_i \log R_i - \sum C_j \log C_j + N \log N\right)$$

donde

$$\sum f_{0ij}\log f_{0ij}$$

es la suma de los valores observados por el logaritmo del valor observado en cada casilla de la tabla de contingencia.

$$\sum R_i \log R_i$$

TABLA 33.7 *Marginales y frecuencia de casillas*

	C_1	C_2	C_3	
R_1	$x_{11} = 34$	$x_{12} = 26$	$x_{13} = 75$	$x_{1+} = 135$
R_2	$x_{21} = 20$	$x_{22} = 30$	$x_{23} = 82$	$x_{2+} = 132$
	$x_{+1} = 54$	$x_{+2} = 56$	$x_{+3} = 157$	$x_{++} = 267$

es la suma del total de renglón por el logaritmo de los totales de renglón,

$$\Sigma C_j \log C_j$$

es la suma de los totales de columna por el logaritmo de los totales de columna. La "N" en $N \log N$ es el conteo de la frecuencia total. El uso de dichos símbolos sólo es efectivo cuando se habla de tablas bidimensionales. Con tablas de tres factores o de factores múltiples, los investigadores usarían una notación diferente. Los valores de casilla observados para una tabla de tres factores se escribirían x_{ijk}. Lo que se designaría como total de renglón y totales de columna en una tabla de dos factores, ahora se llaman *totales marginales*. Para una tabla de tres factores, se anotarían como x_{+jk}, x_{i+k}, x_{ij+}. El gran total o el número total de conteos es x_{+++}. Estas notaciones también son útiles para una tabla de contingencia de dos factores. La tabla 33.7 muestra la relación de los componentes de contingencia y la notación.

Es posible reescribir la ecuación de bondad de ajuste para una tabla de contingencia de dos factores como

$$G^2 = 2(\Sigma x_{ij} \log x_{ij} - \Sigma x_{i+} \log x_{i+} - \Sigma x_{+j} \log x_{+j} + x_{++} \log x_{++})$$

Para los datos del ejemplo en la tabla 33.7,

$$\begin{aligned} G^2 &= 2\,(34 \log 34 + 26 \log 26 + 75 \log 75 + 20 \log 20 + 30 \log 30 \\ &\quad + 82 \log 82 - 135 \log 135 - 132 \log 132 - 54 \log 54 - 56 \log 56 \\ &\quad - 157 \log 157 + 267 \log 267) \\ &= 2\,(119.8996 + 84.7105 + 323.8116 + 59.9147 + 102.0359 + 361.3510 \\ &\quad - 662.2121 - 644.5299 - 215.4051 - 225.4197 - 793.8306 + 1491.7954) \\ &= 2\,(2.1213) = 4.2426 \end{aligned}$$

Si se hubiera calculado χ^2 en lugar de G^2, el valor de χ^2 sería 4.1943. El cálculo de este valor requirió del cálculo adicional de los valores esperados, fe_{ij}. Para este ejemplo serían $fe_{11} = 27.3$, $fe_{12} = 28.31$, $fe_{13} = 79.38$, $fe_{21} = 26.7$, $fe_{22} = 27.69$ y $fe_{23} = 77.62$. Como se puede ver, G^2 y χ^2 no producen el mismo valor. Sin embargo, ambos valores están evaluados con el mismo número de grados de libertad y también con la misma tabla de chi cuadrada.

Ejemplo de investigación

Puesto que Glick *et al.* (1988) presentaron sus datos en su artículo, éstos se utilizarán para ilustrar la aproximación logarítmica lineal de las tablas de contingencia de factores múltiples. Glick *et al.* estudiaron tres variables: *obediencia* (O), *distancia* (D) y *pertenecia al grupo* (G). La variable *obediencia* tenía dos categorías: *obedeció* o *se rehusó*. La *pertenecia al grupo* implicaba si el solicitante de un favor era o no-miembro del mismo grupo que el participante. Esto es, ¿tenía el solicitante una apariencia física similar al participante? Si la respuesta era "sí",

entonces se consideraba un cómplice *dentro del grupo*; si la respuesta era "no", entonces se trataba de un cómplice *fuera del grupo*. La variable *distancia* medía tres distancias: cerca, medio y lejos. Los investigadores hipotetizaron que las personas estarían dispuestas a obedecer si el cómplice *fuera del grupo* estaba más lejos de ellas que los cómplices *dentro del grupo*. Los investigadores suponían básicamente una interacción de tres factores. Se puede escribir el modelo log-lineal de Glick *et al.* como

$$\log_e(m_{ij}) = \mu + \mu_C + \mu_D + \mu_G + \mu_{OD} + \mu_{OG} + \mu_{DG} + \mu_{ODG}$$

Si se puede obtener un ajuste adecuado entre los valores esperados y los valores observados *sin* el término de interacción de los tres factores, ello indicaría que los datos no pudieron justificar la interacción de los tres factores. Siendo así, los datos no apoyarían la hipótesis de los investigadores, lo cual significaría que el efecto de la distancia interpersonal sobre la obediencia no es diferente en los miembros *dentro del grupo* y *fuera del grupo*.

Al utilizar el programa computacional *ILOG* de Bakeman y Robinson (1994) que acompaña su libro de texto, se obtuvieron los resultados presentados en la tabla 33.8.

Aquí se percibe lo que sucedió con la G^2 cuando se ajustó el modelo saturado. El modelo saturado se ajusta perfectamente a los datos observados. Después se elimina el término de interacción de los tres factores del modelo; esto es, el modelo se prueba:

$$\log_e(m_{ij}) = \mu + \mu_O + \mu_D + \mu_G + \mu_{OD} + \mu_{OG} + \mu_{DG}$$

Si el estadístico G^2 no es significativo, se sabe que se ha encontrado por lo menos un modelo que se ajusta bien a los datos observados. Cuando el modelo se ajusta, la G^2 que se obtiene es 12.4 con 2 grados de libertad. Si se consulta una tabla de χ^2 para $\alpha = 0.05$ y $gl = 2$ el valor crítico es 5.99. Puesto que 12.4 es mayor que 5.99, entonces se tiene una prueba χ^2 estadísticamente significativa, y esto indica que el modelo no se ajusta. Al observar la tabla 33.8 se han listado los resultados de la prueba estadística para cada modelo. Todos los modelos reducidos probados son estadísticamente significativos, lo cual indica que los modelos que se probaron no se ajustan a los datos observados. Como el único modelo que se ajusta a los datos es el modelo saturado, entonces se llega a la misma conclusión que encontraron Glick *et al.*: existe una interacción de tres factores entre las variables.

Análisis multivariado e investigación científica

A pesar de que la revisión de los métodos multivariados ha sido más bien superficial, se debe hacer un alto para ubicarlos dentro del esquema de la investigación para evaluarlos. Por ejemplo, ¿se debe abandonar el análisis de varianza sólo porque la regresión múltiple puede lograr todo lo que hace el análisis de varianza, y más? Tal vez implicaciones como éstas ya se han captado por el lector. ¿En realidad el análisis de regresión múltiple no es

TABLA 33.8 *Análisis de los datos de Glick et al.*

Modelo	G^2	*gl*	Sig.	*Término eliminado*	ΔG^2	Δgl
[ODG] (saturado)	0.0	0	$p > 0.05$	—		
[DG][OD][OG]	12.4	2	$p < 0.005$	ODG	12.4	2
[OD][OG]	13.0	4	$p < 0.05$	DG	0.6	2
[OG][D]	20.2	6	$p < 0.005$	OD	7.2	2
[DG][O]	26.8	7	$p < 0.001$	OG	6.6	1

adecuado para los datos experimentales debido a que es uno de los llamados métodos de correlación (lo cual es sólo en parte)? Otras preguntas importantes pueden y deben plantearse y responderse, especialmente en este punto del desarrollo de la investigación de las ciencias del comportamiento. Quizás estemos en un momento de una importante transición. Desde que Fisher inventó y expuso el análisis de varianza en los años veinte y treinta, el método, o más bien el modelo, ha ejercido una gran influencia en la investigación del comportamiento, particularmente en la psicología. ¿Estamos a punto de superar esta etapa? ¿Hemos entrado en una "etapa multivariada"? Si es así, existiría una influencia enormemente importante sobre el tipo y calidad de investigación realizada por psicólogos, sociólogos y educadores en este nuevo siglo. Evidentemente no es posible manejar todas estas preguntas en un libro de texto. Pero al menos se debe abrir la puerta al estudiante.

¿El análisis de varianza debe ser suplantado por el análisis de regresión múltiple? Los autores no creen que deba ser así. ¿Pero es esto una mera atadura sentimental a algo que se ha encontrado interesante y satisfactorio? Tal vez. Pero hay más cosas que hacer que eso. No tiene mucho sentido utilizar la regresión múltiple en la situación de un problema ordinario de análisis de varianza: la asignación aleatoria de los sujetos a los tratamientos experimentales; número de sujetos iguales o proporcionales en las casillas; una, dos o tres variables independientes. Otro argumento para el análisis de varianza es su utilidad en la enseñanza. El análisis de regresión múltiple, aunque elegante y poderoso, carece de la calidad heurística estructural del análisis de varianza. No existe nada tan efectivo en la investigación de la enseñanza y el aprendizaje como el dibujo de paradigmas de los diseños utilizando la partición analítica del análisis de varianza.

La respuesta es que ambos métodos deben enseñarse y aprenderse. Las demandas adicionales sobre el maestro y el estudiante son inevitables, de la misma manera en que el desarrollo, crecimiento y uso de la estadística inferencial anteriormente en el siglo XX hicieron que su enseñanza y aprendizaje resultaran inevitables. No obstante, la regresión múltiple y otros métodos multivariados sin duda sufrirán de falta de comprensión, inclusive alguna oposición, de la misma forma en que la estadística inferencial lo ha sufrido. Aun en la actualidad existen psicólogos, sociólogos y educadores que saben muy poco sobre estadística inferencial o análisis moderno, y quienes inclusive se oponen a su aprendizaje y a su uso. Sin embargo, esto forma parte de la psicología social y patología del tema. A pesar de que sin duda habrá un retraso cultural, la aceptación definitiva de estas poderosas herramientas de análisis probablemente está asegurada.

Los métodos multivariados, como se ha visto, no son tan fáciles de usar y de interpretar como los métodos univariados. Lo anterior se debe no sólo a su complejidad; se debe más bien a la complejidad de los fenómenos con los que trabajan los científicos del comportamiento. Uno de los atrasos de la investigación educativa, por ejemplo, ha sido que la enorme complejidad de una escuela o de un salón de clases no puede manejarse adecuadamente por los métodos demasiado simples que se utilizan. Algunos científicos consideran que ellos nunca pueden captar el mundo "real" con sus métodos de observación y de análisis. Se trata de individuos atados a la simplificación de las situaciones y problemas que estudian. Nunca pueden "ver el todo", de la misma manera que ningún ser humano es capaz de observar y comprender la totalidad de cierto fenómeno. Pero los métodos multivariados captan la realidad psicológica, sociológica y educativa mejor que métodos más simples, y permiten que los investigadores manejen porciones mayores de sus problemas de investigación. En la investigación educativa, los días del experimento de métodos simples con un grupo experimental y un grupo control están contados. En la investigación sociológica, la reducción de gran cantidad de datos valiosos a frecuencias y cruces de porcentajes disminuirá en relación con el cuerpo total de la investigación sociológica.

Lo más importante de todo, el futuro sano de la investigación del comportamiento depende del desarrollo sano de las teorías psicológicas, sociológicas y de otros tipos, para ayudar a explicar las relaciones entre los fenómenos del comportamiento. Por definición, las teorías forman conjuntos interrelacionados de constructos o variables. Evidentemente los métodos multivariados están bien adaptados para comprobar formulaciones teóricas bastante complejas, pues su campo natural es el análisis simultáneo de diversas variables. De hecho, el desarrollo de la teoría del comportamiento debe ir de la mano, e inclusive depender de la asimilación, la maestría y del uso inteligente de los métodos multivariados. En los próximos dos capítulos se ofrecerá una visión multivariada diferente para realizar investigación en las ciencias sociales y del comportamiento.

RESUMEN DE CAPÍTULO

1. Este capítulo examina las diferencias y similitudes entre el análisis de varianza y la regresión múltiple.
2. La regresión múltiple puede, en esencia, hacer todos los análisis que el ANOVA, y aún más.
3. El análisis de varianza posee una estructura que es intuitivamente atractiva para los investigadores.
4. En el presente capítulo se analizan las diferencias entre los diferentes tipos de codificación: dummy, de efectos y ortogonal. Cada una produce la misma R^2, pero los coeficientes individuales de las variables son diferentes.
5. La diferencia entre el análisis de varianza y el análisis de covarianza reside en que en el ANOVA las variables independientes no están correlacionadas. En el análisis de covarianza, por lo menos una variable está correlacionada con las otras variables independientes.
6. Esta variable correlacionada se llama *covariable*. Sirve para eliminar su varianza de la variable dependiente, antes de que se prueben las variables independientes.
7. El análisis de covarianza se maneja con facilidad por medio de la regresión múltiple.
8. El análisis discriminante resulta similar a la regresión múltiple, con algunas excepciones. En el análisis discriminante la variable dependiente es categórica. Además ofrece un estadístico que indica qué tan bien la función discriminante clasifica observaciones.
9. En la correlación canónica, en lugar de una variable dependiente como en la regresión múltiple y en el análisis discriminante, existen más variables dependientes. El objetivo es encontrar dos conjuntos de coeficientes que maximicen la varianza entre los dos conjuntos de variables.
10. El análisis multivariado de varianza o MANOVA es el equivalente multivariado del análisis de varianza. En el ANOVA univariado el análisis se hace para una variable dependiente a la vez. En el análisis multivariado de varianza, se consideran al mismo tiempo múltiples variables dependientes.
11. Los MANOVA, al igual que todos los métodos multivariados, pueden llevar a resultados que sean difíciles de interpretar.
12. El análisis de ruta utiliza los pesos estandarizados de la regresión para estudiar los efectos directos e indirectos de unas variables sobre otras. Se utiliza mejor como un modelo conceptual a probarse.
13. El análisis de ruta implica el dibujo de un diagrama de ruta que muestre cómo se relacionan las variables.

14. La regresión de cresta fue utilizada primero en ingeniería química por Arthur Hoerl. Posteriormente se convirtió en una herramienta para otras áreas.
15. Los mínimos cuadrados ordinarios son el método estadístico utilizado por la mayoría de los programas computacionales de regresión múltiple. No obstante, cuando las variables independientes están altamente correlacionadas entre sí, los mínimos cuadrados presentan problemas en términos de estimación.
16. La regresión de cresta agrega un sesgo a la ecuación y, por ende, estabiliza los coeficientes de regresión. La regresión de cresta constituye un tema polémico en las ciencias del comportamiento.
17. La regresión logística es el modelo popular y alternativo al análisis discriminante. No tiene tantas restricciones como las impuestas al análisis discriminante.
18. Con menos restricciones, la regresión logística puede manejar una diversidad de problemas. Tal como el análisis discriminante, la regresión logística posee una variable dependiente categórica.
19. Las tablas de contingencia de factores múltiples se manejan utilizando el análisis log-lineal.
20. La idea principal que está detrás del análisis log-lineal para las tablas de contingencia de factores múltiples es encontrar el modelo apropiado que explicará la variación de los valores observados.
21. Existen modelos jerárquicos y no jerárquicos en el análisis log-lineal. El jerárquico es el más útil. Los modelos no jerárquicos están sujetos a problemas de interpretación.

SUGERENCIAS DE ESTUDIO

1. Por desgracia son escasos los tratamientos elementales de regresión múltiple completamente satisfactorios, en especial si se espera un tratamiento de regresión concomitante al análisis de varianza. Tal vez no sea posible un tratamiento elemental satisfactorio de un tema tan complejo. Las siguientes referencias sobre regresión múltiple y otros métodos multivariados pueden resultar de utilidad. Algunas de ellas también se incluyen en la sección de referencias, debido a que se citaron en el presente capítulo.

Kerlinger, F. y Pedhazur, E. (1973). *Multiple regression in behavioral research*. Nueva York: Holt, Rinehart and Winston. [Un texto que intenta incrementar la comprensión de la regresión múltiple y sus usos en la investigación por medio de la presentación de una exposición lo más simple posible, así como diversos ejemplos con números sencillos. También incluye un programa computacional completo de regresión múltiple en el apéndice.]

Pedhazur, E. (1996). *Multiple regression in behavioral research: Explanation and prediction* (4a. ed.). Orlando, Florida: Harcourt Brace. [Es la revisión del libro de Kerlinger y Pedhazur. Sin embargo, es más detallado y profundo. Bastante recomendable.]

Stevens, J. P. (1996). *Applied multivariate statistics for the social sciences* (3a. ed.). Mahwah, Nueva Jersey: Lawrence Erlbaum. [Un libro sobre estadística multivariada fácil de leer. Incluye notas respecto a los resultados en computadora de conocidos programas estadísticos.]

Tabachnick, B. y Fidell, L. (1996). *Using multivariate statistics* (3a. ed.). Nueva York: HarperCollins. [Muestra la estadística multivariada desde un punto de vista de

resultados computacionales. Muy útil para quienes desean aprender estadística multivariada y también sobre los programas computacionales utilizados para realizar los análisis.]

Una vez que el estudiante y el investigador manejan los elementos del análisis de regresión múltiple y que han tenido alguna experiencia con problemas reales, las siguientes referencias proporcionan una guía sofisticada en el uso del análisis de regresión múltiple y, más importante aún, en la interpretación de los datos.

Cohen, J. y Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences* (2a. ed.). Mahwah, Nueva Jersey: Lawrence Erlbaum. [Un tratamiento excelente de la regresión múltiple. Muestra cuantos problemas donde se utilizó el análisis de varianza, podrían haberse analizado por medio de regresión múltiple. También muestra la forma en que la regresión múltiple se utiliza para estudiar causalidad.]

Daniel, C. Wood, F. S. y Gorman, J. W. (1980). *Fitting equations to data: Computer analysis of multifactor data* (2a. ed.). Nueva York: Wiley. [Resume, por medio de ejemplos, la forma en que estos estadísticos se aproximaron al análisis de datos de investigación, donde los investigadores no siguieron los requisitos estándares del diseño estadístico de experimentos. Utiliza muchos análisis y explicaciones generados por computadora.]

Draper, N. y Smith, H. (1981). *Applied regression analysis* (2a. ed.). Nueva York: John Wiley & Sons. [Un clásico en el campo del análisis de regresión. Requiere de alguna sofisticación matemática, pero es un libro útil y citado con frecuencia.]

Kleinbaum, D. G., Kupper, L. L., Muller, K. E. y Nizam, A. (1997). *Applied regression analysis and other multivariate methods.* (3a. ed.). Belmont, California: Duxbury. [Esta obra aclara la confusión entre multivariado y multivariable. Se mencionó brevemente en el capítulo 2. Vale la pena leerlo.]

Mendenhall, W. (1968). *An introduction to linear models and the design and analysis of experiments.* Belmont, California: Wadsworth. [Sin duda uno de los mejores libros escritos, que presenta, sin sofisticación, el uso de la regresión múltiple (modelo lineal general) en lugar del análisis de varianza. Requiere de conocimiento sobre álgebra de matrices. Se trata de un libro que ya no se imprime.]

Neter, J. Wasserman, W. y Kutner, M. H. (1996). *Applied linear regression models* (3a. ed.). Burr Ridge, Illinois: Irwin. [Este libro es similar al escrito por Woodward, Bonett y Brecht, citado más adelante. Cubre bien el uso de la regresión. Requiere de conocimiento sobre álgebra de matrices.]

Los siguientes textos son fundamentales: enfatizan las bases teóricas y matemáticas de los métodos multivariados.

Carroll, J. D. y Green, P. (1997). *Mathematical tools for applied multivariate analysis* (3a. ed.). Nueva York: Academic Press. [Un libro sobresaliente sobre las bases matemáticas del análisis multivariado. Altamente recomendable.]

Kenny, D. (1979). *Correlation and causality.* Nueva York: John Wiley & Sons. [Merece muchas horas de estudio.]

Wickens, T. D. (1994). *The geometry of multivariate statistics.* Mahwah, Nueva Jersey: Lawrence Erlbaum. [Presenta los procedimientos de la estadística multivariada de forma geométrica. Ayuda al estudiante a conceptualizar las relaciones multivariadas.]

2. Suponga que un psicólogo social tiene dos matrices de correlación:

$$\begin{array}{c} \\ X_1 \\ X_2 \\ Y \end{array}\begin{array}{c} \begin{array}{ccc} X_1 & X_2 & Y \end{array} \\ \begin{bmatrix} 1.00 & 0 & .70 \\ 0 & 1.00 & .60 \\ .70 & .60 & 1.00 \end{bmatrix} \end{array} \quad \text{A} \qquad \begin{array}{c} \\ X_1 \\ X_2 \\ Y \end{array}\begin{array}{c} \begin{array}{ccc} X_1 & X_2 & Y \end{array} \\ \begin{bmatrix} 1.00 & .40 & .70 \\ .40 & 1.00 & .60 \\ .70 & .60 & 1.00 \end{bmatrix} \end{array} \quad \text{B}$$

a) ¿Cuál matriz, la A o la B, producirá la R^2 mayor? ¿Por qué?
b) Calcule la R^2 de la matriz A.
[Respuestas: *a)* la matriz A; *b)* $R^2 = .85$.]

3. A continuación se presentan tres conjuntos de datos ficticios simples, ordenados para un análisis de varianza. Ordene los datos para un análisis de regresión múltiple y realice tanto del análisis de regresión como le sea posible. Utilice codificación dummy (1, 0), como en la tabla 33.2. Los coeficientes b son: $b_1 = 3$; $b_2 = 6$.

A_1	A_2	A_3
7	12	5
6	9	2
5	10	6
9	8	3
8	11	4

Imagine que A_1, A_2 y A_3 son tres métodos para cambiar las actitudes raciales y que la variable dependiente es una medida del cambio en donde las puntuaciones más altas indican mayor cambio. Interprete los resultados. [Respuestas: a = 4; R^2 = .75; F = 18, con gl = 2, 12; $sc_{reg} = 90$; $sc_t = 120$. Observe que estos datos ficticios son en realidad las puntuaciones de la tabla 33.2 con un 1 que se agrega a cada puntuación. Compare los diversos estadísticos de regresión y de análisis de varianza anteriores, con aquellos calculados con los datos de la tabla 33.2.]

4. Utilizando los datos de la tabla 34.2 en el capítulo 34, calcule las sumas de cada par de X_1 y X_2. Correlacione dichas sumas con las puntuaciones de Y. Compare el cuadrado de esta correlación con $R^2_{y.12} = .51$ ($r^2 = .70^2 = .49$). Puesto que los valores son bastante cercanos, ¿por qué no debemos simplemente utilizar los promedios de las variables independientes sin molestarnos con la complejidad del análisis de regresión múltiple?

5. A continuación se presenta una lista de varios estudios interesantes que han utilizado regresión múltiple, análisis de ruta y análisis discriminante de forma efectiva. Lea uno o dos de ellos cuidadosamente.

Abel, M. H. (1998). Interaction of humor and gender in moderating relationships between stress and outcomes. *Journal of Psychology, 132*, 267-276. [Utiliza la regresión múltiple para estudiar los efectos moderadores del humor sobre el estrés y la ansiedad.]

Bachman, I. y O' Malley, P. (1977). Self-esteem in young men: A longitudinal analysis of the impact of educational and occupational attainment. *Journal of Personality*

and Social Psychology, 35, 365-380. [Un estudio educativo sobresaliente que utiliza el análisis de ruta. Los resultados son contrarios a lo esperado.]

Fischer, C. (1975). The city and political psychology. *American Political Science Review, 69*, 559-571. [Utiliza el análisis de ruta para estudiar el sentido de eficacia política.]

Frederick, C. M. y Morrison, C. S. (1998). A mediational model of social physique anxiety and eating disordered behaviors. *Perceptual and Motor Skills, 86*, 139-145. [Desarrolla un modelo de ruta muy simple y fácil de entender, en relación con la ansiedad sobre el físico, los rasgos del trastorno de alimentación y el comportamiento del trastorno de alimentación.]

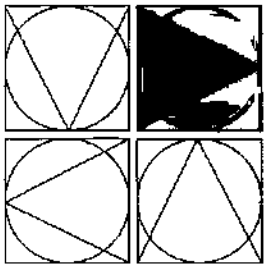
Leith, K. P. y Baumeister, R. F. (1998). Empathy, shame, guilt and narratives of interpersonal conflicts: Guilt prone people are better at perspective taking. *Journal of Personality, 66*, 11-39. [Utiliza análisis multivariado de varianza, análisis de covarianza y análisis de ruta para estudiar a personas con tendencia a sentirse culpables.]

Marjoribanks, K. (1972). Ethnic and environmental influences on mental abilities. *American Journal of Sociology, 78*, 323-337. [Un uso interesante de la suma y resta de las R^2 para evaluar la influencia relativa de variables, especialmente del ambiente y la etnicidad.]

Onwuegbuzie, A. J. (1997). The teacher as researcher: The relationship between research anxiety and learning style in a research methodology course. *College Student Journal, 31*, 496-506. [Este estudio utiliza la regresión múltiple para determinar algunas de las características de los maestros ansiosos por la investigación, en términos del tipo de estilo de aprendizaje.]

Ronis, D. L., Antonakos, C. L. y Lang, W. P. (1996). Usefulness of multiple equations for predicting preventive oral health behaviors. *Health Education Quarterly, 23*, 512-527. [Analiza los resultados de un estudio que utiliza la correlación canónica. Los investigadores encontraron tres funciones que explicaban tres conductas de salud oral.]

Vincke, J. y Bolton, R. (1997). Beyond the sexual model: Combining complementary cognitions to explain and predict unsafe sex among gay men. *Human Organization, 56*, 38-46. [Emplea tanto el análisis multivariado de varianza como el análisis discriminante para evaluar los placeres y peligros de las prácticas sexuales sin protección.]



CAPÍTULO 34

ANÁLISIS FACTORIAL

Muchos investigadores consideran el análisis factorial como la reina de los métodos analíticos, debido a su poder, elegancia y cercanía al corazón del propósito científico. Sin embargo, se trata de un método que no está libre de controversia. A pesar de que se trata de un método poderoso, no constituye una panacea para estudios mal diseñados o sin diseño. Comrey (1978) señaló que el análisis factorial ha sido un tema de gran discusión y crítica. No obstante, a pesar de las críticas, el aumento de su uso continúa. En este capítulo se explorará lo que es el análisis factorial, y por qué y cómo se realiza. También se explorarán las dificultades que un investigador llega a encontrar si no tiene cuidado al utilizar este poderoso método. Durante la exploración también se examinará investigación pasada y actual, donde el análisis factorial ha sido la metodología central.

El análisis factorial sirve a la causa de la parsimonia científica. Reduce la multiplicidad de las pruebas y medidas a una mayor simplicidad. En efecto, indica qué pruebas o medi-

das van juntas —las que virtualmente miden lo mismo— y qué tanto es así. Por lo tanto, reduce el número de variables con las que el científico debe enfrentarse. También ayuda al científico a ubicar e identificar unidades o propiedades fundamentales que subyacen a pruebas y medidas. Suponga que un investigador ha medido 20 variables en un grupo de personas. Se calculan las correlaciones entre las variables y se sintetizan en forma de una matriz de correlación. Cuando un investigador examina una tabla de correlación entre variables, resulta muy difícil interpretar lo que en realidad está sucediendo. Por lo general, es muy difícil encontrar un patrón de correlaciones interpretable. El análisis factorial está diseñado para tomar esas correlaciones y encontrar algún orden entre ellas. El método está diseñado para encontrar lo que las variables tienen en común. Aun cuando en el presente capítulo la explicación sobre el análisis factorial se centrará en el uso de los coeficientes de correlación, el análisis factorial no se limita tan sólo a matrices de correlación. Sin embargo, en las ciencias sociales, del comportamiento y educativas, la correlación es el índice utilizado con mayor frecuencia en un análisis factorial.

Un *factor* es un constructo, una entidad hipotética, una variable latente que se asume como el fundamento de pruebas, escalas, reactivos y, de hecho, de medidas de casi cualquier clase. Para aquellos investigadores en ciencias del comportamiento que estén desarrollando escalas o pruebas, el análisis factorial les sirve para ofrecer evidencia de la ausencia o presencia de validez. Se ha encontrado una serie de factores que son fundamento de la inteligencia, por ejemplo: destreza verbal, habilidad numérica, razonamiento abstracto, razonamiento espacial, memoria, etcétera. De forma similar, también se han aislado e identificado factores de aptitud, actitud y personalidad. ¡Inclusive se han encontrado factores de naciones y personas!

Fundamentos

Breve historia

El desarrollo del método denominado análisis factorial se atribuye a Charles Spearman. En 1904, Spearman publicó un artículo de 93 páginas que cubría su teoría sobre la inteligencia y el desarrollo del método para confirmar su teoría de que un factor común explicaba toda la inteligencia humana. Ese único factor se denomina *g*, o factor general. Spearman analizó tablas de correlación entre pruebas psicológicas y demostró que había un factor común a todas las pruebas. Las varianzas restantes se atribuían a la prueba específica. Algunas ocasiones la teoría de Spearman se conoce como la teoría de los dos factores. Spearman también vinculó su teoría con la neuropsicología al afirmar que el factor *g* abarcaba toda la corteza humana o sistema nervioso. Muchos investigadores realizaron diversos trabajos respecto a la teoría de Spearman. Algunos de los nombres más notables se mencionan en Ferguson (1971) o Carroll (1993). El concepto de *g* resulta polémico. A pesar de que muchos autores han desarrollado estudios empíricos para demostrar que no existe, su uso y referencia permanecen. Uno de los principales antagonistas de la teoría de Spearman fue L. L. Thurstone (1947), quien proporcionó evidencia inicial de que había factores de inteligencia, a los que llamó "habilidades mentales primarias". Aunque Thurstone proporcionó evidencia en contra de Spearman, más tarde se demostró que su teoría de inteligencia también era cuestionable.

A pesar de tales raíces, comienzos e intenciones iniciales del análisis factorial, los métodos creados como resultado del trabajo de Spearman y de Thurstone continúan siendo importantes. En sus esfuerzos por "probar" que el otro estaba equivocado, surgieron

los métodos modernos y creativos del análisis factorial. Los investigadores continúan utilizando dichos métodos hoy en día. La contribución de Thurstone al campo es monumental. La mayor parte del análisis factorial es el resultado directo de su trabajo. El profesor Andrew Comrey de la UCLA le dijo en una ocasión al segundo autor de este libro, que Thurstone (1947) estableció las bases para cerca del 90 por ciento del análisis factorial moderno. Desde entonces, los investigadores intentan definir el restante 10 por ciento. Thurstone tomó de manera efectiva el método del análisis factorial de Spearman y lo mejoró. Thurstone fue responsable del desarrollo del método centroide. Antes del advenimiento de la computadora de alta velocidad, los análisis factoriales se realizaban a mano. El método centroide ofreció una muy buena aproximación al método más poderoso desarrollado por Hotelling (1933). El método del factor principal de Hotelling se adaptaba mejor para las computadoras; aunque era bastante laborioso para ser realizado mediante cálculos manuales. Con el propósito de facilitar la interpretación de los resultados analíticos del factor, Thurstone también desarrolló el método de rotación y el concepto de estructura simple. La estructura simple es uno de los desarrollos clave dentro de la metodología analítica del factor.

En este capítulo y en el siguiente se procurará hacer las menores referencias posibles al álgebra matricial. Sin embargo, la explicación del análisis factorial requiere considerar las matrices. Algunas explicaciones se facilitan al utilizarlas, en especial cuando se exponga el análisis factorial confirmatorio y el modelamiento de la ecuación estructural.

Un ejemplo hipotético

Considere que se aplican seis pruebas a un gran número de alumnos de primero de secundaria. Se sospecha que las seis pruebas están midiendo no seis, sino un menor número de variables. Las pruebas son *vocabulario, lectura, sinónimos, números, aritmética* (prueba estandarizada), *aritmética* (prueba elaborada por el profesor). Los nombres de estas pruebas indican su naturaleza. Se les denomina, respectivamente, *V, L, S, N, AE, AEP.* (Aunque las últimas dos pruebas son de aritmética, tienen un contenido diferente. Se asume que existe una buena razón para incluir ambas en la pequeña batería de pruebas.) Una vez que las pruebas se aplican y se evalúan, se calculan los coeficientes de correlación de cada prueba con todas las demás. Se ordenan las r en una matriz de correlación (comúnmente llamada matriz R). La matriz se presenta en la tabla 34.1.

Recuerde que una matriz es cualquier arreglo rectangular de números (o símbolos). Las matrices de correlación siempre son cuadradas y simétricas; ello se debe a que la mitad que se encuentra por debajo de la diagonal principal (de la parte izquierda superior a la

TABLA 34.1 *Matriz R: coeficientes de correlación entre seis pruebas*

		V	*L*	*S*	*N*	*AE*	*AEP*
	V		.72	.63	.09	.09	.00
Conglomerado I	*L*	.72		.57	.15	.16	.09
	S	.63	.57		.14	.15	.09
	N	.09	.15	.14		.57	.63
	AE	.09	.16	.15	.57		.72
	AEP	.00	.09	.09	.63	.72	

Conglomerado II

parte inferior derecha) es igual a la mitad superior de la matriz. Es decir, los coeficientes de la mitad inferior son idénticos a los de la mitad superior, con excepción de su arreglo. (Note que el renglón superior es igual a la primera columna, el segundo renglón es igual a la segunda columna y así sucesivamente). Si se intercambian los renglones y las columnas de la matriz de correlación, la matriz resultante será idéntica a la matriz original. Cuando así sucede, se sabe que la matriz es simétrica. También, cuando se intercambian los renglones con las columnas, la matriz resultante se denomina de *transposición*. Si se tiene una matriz llamada **A**, la de transposición se llama $\mathbf{A}^{T}$. Este concepto se utilizará más adelante.

El problema al que hay que enfrentarse se expresa en dos preguntas: ¿cuántas variables o factores subyacentes existen? ¿Cuáles son los factores? Se presume que son unidades subyacentes a los desempeños de las pruebas, que se reflejan en los coeficientes de correlación. Si dos o más pruebas están altamente correlacionadas, entonces las pruebas comparten varianza; tienen varianza de factor común, y están midiendo algo en común.

La primera pregunta, en este caso, es fácil de contestar. Existen dos factores, lo cual está indicado por los dos grupos de r circuladas y llamadas *conglomerado I* y *conglomerado II* en la tabla 34.1. Observe que *V* se correlaciona con *L* en .72; *V* con *S* en .63; y *L* con *S* en .57. *V*, *L* y *S* parecen medir algo en común. De manera similar, *N* se correlaciona con *AE* en .57 y con *AEP* en .63; y *AE* se correlaciona con *AEP* en .72. *N*, *AE* y *AEP* miden algo en común. Las pruebas en el *conglomerado I*, aunque correlacionadas entre sí, no están muy correlacionadas con las pruebas en el *conglomerado II*. De la misma manera, aunque *N*, *AE* y *AEP* se correlacionan entre sí, no están altamente correlacionadas con las pruebas *V*, *L* y *S*. En efecto, lo que miden en común las pruebas del *conglomerado I*, no es lo mismo que miden en común las pruebas del *conglomerado II*. Parece haber dos conglomerados o factores en la matriz. El lector debe notar que en esta presentación se utilizan sobresimplificaciones ocasionales y ejemplos poco realistas. La matriz **R** de la tabla 34.1 no es realista. Todas las pruebas estarían correlacionadas positivamente, y quizá los dos factores surgirían. Además, aunque los grupos asemejan factores, *no* son factores. Sin embargo, por simplicidad y por razones pedagógicas, se arriesgó a efectuar tales sobresimplificaciones.

Al estudiar la matriz **R** se determina que existen dos factores detrás de estas pruebas. Respecto a la segunda pregunta (¿cuáles son los factores?) casi siempre es más difícil responder. Cuando se pregunta cuáles son los factores, se busca nombrarlos. Se buscan *constructos* que expliquen las unidades subyacentes o las varianzas de factor común de los factores. Se pregunta qué es lo que tienen en común las pruebas *V*, *L* y *S*, por un lado, y las pruebas *N*, *AE* y *AEP*, por el otro. *V*, *L* y *S* son pruebas de vocabulario, lectura y sinónimos. Las tres involucran palabras, en un sentido amplio. Quizás el factor subyacente sea *habilidad verbal*. Se denomina al factor *verbal* o *V*. Las pruebas *N*, *AE* y *AEP* involucran operaciones numéricas o aritméticas. Suponga que se denomina a este factor *aritmética*. Un amigo señala que la prueba *N* en realidad no involucra operaciones aritméticas, pues consiste principalmente de la manipulación no aritmética de números, lo cual se pasó por alto a causa de la insistencia por dar un nombre a la unidad subyacente. De cualquier manera, ahora se llama al factor *numérico* o *N*. No existe ninguna inconsistencia: las tres pruebas involucran números, manipulación y operación numérica.

Ya se respondieron las dos preguntas: existen dos factores y se denominan *verbal*, *V*, y *numérico*, *N*. No obstante, debe señalarse rápida y urgentemente que ninguna de las preguntas se contesta por completo en la verdadera investigación analítica factorial. Lo anterior sucede en especial en las investigaciones iniciales en un campo. El número de factores puede cambiar en investigaciones subsecuentes, utilizando las mismas pruebas. Una de las pruebas de *V* puede tener también alguna varianza en común con otro factor, por ejemplo, *K*. Si una prueba que mide *K* se añade a la matriz, quizá surja un nuevo factor. Tal vez más importante, el nombre de un factor puede ser incorrecto. Investigación subsecuente utili-

zando estas pruebas *V* y otras pruebas, demuestra que *V* ahora ya no es común a todas las pruebas. Entonces, el investigador debe encontrar otro constructo, otra fuente de varianza del factor común. En síntesis, los nombres de los factores son tentativos; son hipótesis a comprobarse en análisis factoriales posteriores y en otros tipos de investigación.

Matrices factoriales y cargas factoriales

Si una prueba mide sólo un factor, se dice que es *factorialmente pura*. Dependiendo del grado en que una prueba mida un factor, se dice que está *cargada* o *saturada* con el factor. En realidad el análisis factorial no está completo a menos que se conozca si una prueba es factorialmente pura o qué tan saturada está con un factor. Si una medida no es factorialmente pura, por lo general se busca saber qué otros factores la conforman. Algunas medidas son tan complejas que es difícil decir exactamente qué miden. Un buen ejemplo son las calificaciones del profesor, o los promedios de calificaciones, que pueden consistir de un número de dimensiones del desempeño del estudiante. Si una prueba contiene más de un factor, se dice que es *factorialmente compleja*.

Algunas pruebas y medidas son factorialmente muy complejas. La prueba de inteligencia de Stanford-Binet, la prueba de inteligencia de Otis y la escala *F* (de autoritarismo) son algunos ejemplos. Un hecho apreciado de la investigación científica es tener medidas puras de las variables. Si una medida de habilidad numérica no es factorialmente pura, entonces, ¿cómo se puede confiar en que una relación entre habilidad numérica y rendimiento académico, por ejemplo, es en realidad la relación que se piensa que es? Si la prueba mide tanto habilidad numérica como razonamiento verbal, las relaciones que se estudian con su ayuda resultarán dudosas.

Para resolver éstos y otros problemas se requiere de un método objetivo que determine el número de factores, las pruebas que pesan en los diversos factores y la magnitud de las cargas. Existen varios métodos analíticos factoriales para lograr tales propósitos. Posteriormente se analizará uno de ellos.

Uno de los resultados finales de un análisis factorial es la llamada *matriz factorial*, que es una tabla de coeficientes que expresa la relación entre las pruebas y los factores subyacentes. En la tabla 34.2 se presenta la matriz factorial producida por el análisis factorial de los datos de la tabla 34.1, con el método de factores principales, uno de los diversos métodos disponibles, y con la subsecuente rotación factorial (que se explicará más adelante). Los datos de la tabla se denominan pesos o *cargas factoriales*. Pueden escribirse como a_{ij}, lo que significa la carga *a* de la *prueba i* sobre el *factor j*. En la segunda línea, .79 es la carga

TABLA 34.2 *Matriz factorial de los datos de la tabla 34.1, solución rotada*[a]

Pruebas	*A*	*B*	h^2
V	*.83*	.01	.70
L	*.79*	.10	.63
S	*.70*	.10	.50
N	.10	*.70*	.50
AE	.10	*.79*	.63
AEP	.01	*.83*	.70

[a] Véase el texto para identificar las pruebas. Las cargas "significativas" están en *itálicas*. Véase también los pies de la tabla 34.5.

factorial de la prueba *R* sobre el factor *A*. Algunos de los analistas de factores llaman a los factores de la solución final I, II, ..., o I′, II′, etcétera. En este capítulo se denominan I, II, ..., a los factores sin rotación; y *A*, *B*,..., a los factores con rotación (solución final). En la cuarta línea, .70 es la carga factorial de la prueba *N* en el factor *B*. La prueba *AE* tiene las siguientes cargas: .10 en el factor *A* y .79 en el factor *B*.

Las cargas factoriales no son difíciles de interpretar. Oscilan entre −1.00 y +1.00, como los coeficientes de correlación. Además se interpretan de manera similar. De hecho, expresan las correlaciones *entre las pruebas y los factores*. Por ejemplo, la prueba *V* tiene las siguientes correlaciones con los factores *A* y *B*, respectivamente: .83 y .01. Evidentemente, la prueba *V* tiene una fuerte carga en *A*, pero ninguna en *B*. Las pruebas *V*, *L* y *S* pesan en *A* pero no en *B*. Las pruebas *N*, *AE* y *AEM* pesan en *B* pero no en *A*. Todas las pruebas parecen ser "puras".

Las cifras de la última columna se llaman *comunalidades* o h^2. Son las sumas de cuadrados de las cargas factoriales de una prueba o variable. Por ejemplo, la comunalidad de la prueba *R* es $(.79)^2 + (.10)^2 = .63$. La comunalidad de una prueba o variable es su varianza del factor común, lo cual se explicará más adelante cuando se exponga la teoría de factores.

Antes de continuar, debe señalarse nuevamente que este ejemplo no es realista. Las matrices de factores rara vez presentan una imagen tan clara. De hecho, la matriz factorial de la tabla 34.2 ya era "conocida". El autor primero escribió la matriz presentada en la tabla 34.3. Si esta matriz se multiplica por sí misma, entonces se obtiene la matriz de la tabla 34.1 (con valores en la diagonal). En tal caso, lo que se necesita para obtener **R** es multiplicar cada renglón por cada uno de los otros renglones. Por ejemplo, se multiplica el renglón *V* por el renglón *L*: (.90)(.80) + (.00)(.10) = .72; el renglón *V* por el renglón *S*: (.90)(.70) + (.00)(.10) = .63; el renglón *S* por el renglón *AE*: (.70)(.10) + (.10)(.80) = .15; etcétera. Después, la matriz **R** resultante se analizó factorialmente. Esta operación de multiplicación de la matriz surge de la llamada *ecuación básica del análisis factorial:* $\mathbf{R} = \mathbf{F}\mathbf{F}^T$, que indica, de forma sucinta y en símbolos de matriz, lo que se expuso de forma más elaborada anteriormente. Algunas veces dicha ecuación fundamental se escribe $\mathbf{R} = \mathbf{A}\mathbf{A}^T$ o $\mathbf{R} = \mathbf{P}\Phi\mathbf{P}^T + \mathbf{U}$. La última ecuación es la más general de las tres. Un conocimiento profundo del análisis factorial requiere de un buen entendimiento del álgebra matricial.

Resulta instructivo comparar la tabla 34.2 con la tabla 34.3. Observe las discrepancias, que son pequeñas. Es decir, el método analítico factorial falible no puede reproducir de forma perfecta la "verdadera" matriz factorial; sólo la estima. En tal caso el ajuste es cercano debido a la simplicidad deliberada del problema. Los datos reales no son tan complacientes. Además, nunca se conoce la matriz factorial "verdadera". Si así fuera, no habría necesidad de realizar el análisis factorial. Por lo general, se estima la matriz factorial a partir de la matriz de correlación. La complejidad y falibilidad de los datos de investigación con frecuencia hacen que dicha estimación sea un asunto difícil.

TABLA 34.3 *Matriz factorial original de la cual se derivó la matriz R de la tabla 34.1*

Pruebas	A	B	h^2
V	*.90*	.00	.81
L	*.80*	.10	*.65*
S	*.70*	.10	.50
N	.10	*.70*	*.50*
AE	.10	*.80*	.65
AEP	.00	*.90*	.81

Un poco de teoría factorial

En el capítulo 28 se escribió una ecuación que expresa las fuentes de varianza en una medida (o prueba):

$$V_t = V_{co} + V_{esp} + V_e \tag{34.1}$$

donde V_t es igual a la varianza total de una medida; V_{co} significa la varianza del factor común o la varianza que dos o más medidas comparten en común; V_{esp} equivale a la varianza específica o la varianza de la medida que no es compartida por cualquier otra medida, es decir, la varianza de esa medida y de ninguna otra; V_e es igual a la varianza del error.

La varianza del factor común V_{co} se dividió en dos fuentes de varianza, A y B, que son dos factores (véase ecuación 28.11):

$$V_{co} = V_A + V_B \tag{34.2}$$

V_A podría ser la varianza de la habilidad verbal, y V_B podría ser la varianza de la habilidad numérica, lo cual es razonable si se piensa en las sumas de cuadrados de las cargas factoriales de cualquier prueba:

$$h_i^2 = a_i^2 + b_i^2 + \ldots + k_i^2 \tag{34.3}$$

donde $a_i^2, b_i^2, \ldots$ son los cuadrados de las cargas factoriales de la prueba i, y h_i^2 es la comunidad de la prueba i. Pero $h_i^2 = V_{co}$. Por lo tanto, $V(A) = a^2$ y $V(B) = b^2$, y la ecuación 34.2 se vincula con operaciones analíticas factoriales reales.

Sin embargo, por supuesto, quizás haya más de dos factores. La ecuación generalizada es

$$V_{co} = V_A + V_B + \ldots + V_k \tag{34.4}$$

Sustituyendo en la ecuación 34.1 se obtiene

$$V_t = V_A + V_B + \ldots + V_k + V_{esp} + V_e \tag{34.5}$$

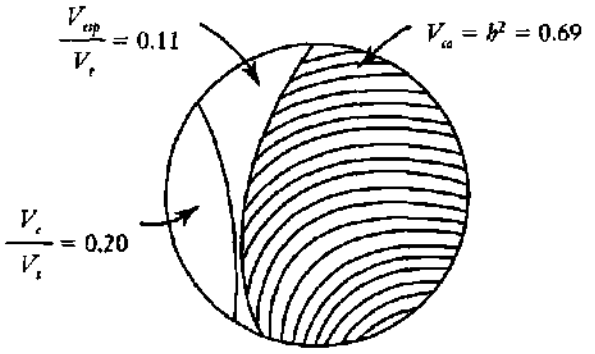
Dividiendo entre V_t, se encuentra la representación proporcional:

$$\underbrace{\overbrace{\frac{V_t}{V_t} = 1.00 = \frac{V_A}{V_t} + \frac{V_B}{V_t}}^{h^2} + \ldots + \frac{V_k}{V_t} + \frac{V_{esp}}{V_t}}_{r_{tt}} + \frac{V_e}{V_t} \tag{34.6}$$

Las partes h^2 y r_{tt} de la ecuación se han denominado de la misma manera que en el capítulo 28. Esta ecuación tiene belleza. Une fuertemente la teoría de medición y la teoría de factores. h^2 es la proporción de la varianza total que es varianza del factor común. r_{tt} es la proporción de la varianza total que es varianza confiable. V_e/V_t es la proporción de la varianza total que es varianza del error. En el capítulo 28 una ecuación como ésta permitió ligar la confiabilidad y la validez. Ahora demuestra la relación entre la teoría de factores y la teoría de medición. Se observa, brevemente, que el *principal problema del análisis factorial consiste en determinar los componentes de varianza de la varianza del factor común total.*

Considere la prueba V en la tabla 34.2. Un vistazo a la ecuación 34.6 indica, entre otras cuestiones, que la confiabilidad de una medida siempre es mayor que, o igual que su comunidad. Entonces, la confiabilidad de la prueba V es, por lo menos, .70. Suponga que $r_{tt} = .80$. Como $V_t/V_t = 1.00$, se pueden especificar todos los términos:

FIGURA 34.1



$$\frac{V_t}{V_t} = 1.00 = \overbrace{\underbrace{(.83)^2 + (.01)^2}_{r_{11}=.80}}^{h^2=.69} + \overbrace{.11}^{V_{sp}} + \overbrace{.20}^{V_e}$$

Entonces, la prueba *V* posee una alta proporción de varianza de factor común y una baja proporción de varianza específica.

Las proporciones se ven claramente en un diagrama circular. Sea el área del círculo de la figura 34.1 igual a la varianza total o 1.00 (100 por ciento del área). Las tres varianzas se han indicado al separar las áreas del círculo. V_{co} o h^2, por ejemplo, es el 69 por ciento, V_{sp} es el 11 por ciento y V_e es el 20 por ciento de la varianza total.

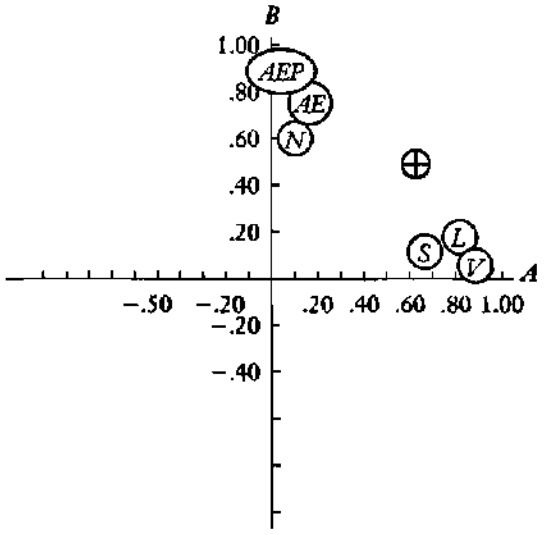
Una investigación analítica factorial que incluya a *V* informaría principalmente sobre V_{co}, la varianza de factor común. Indicaría la proporción de la varianza total de la prueba que es varianza del factor común y ofrecería indicios sobre su naturaleza al indicar qué otras pruebas comparten la misma varianza del factor común, y cuáles no.

Representación gráfica de factores y cargas factoriales

El estudiante del análisis factorial debe aprender a pensar de forma espacial y geométrica, para captar la naturaleza esencial del modelo factorial. Existen varias formas adecuadas para lograrlo. Una tabla de correlaciones se representa por medio del uso de vectores y de los ángulos entre ellos. Aquí se utiliza un método más común. Se trata a las cifras en los renglones de la matriz factorial como coordenadas y se les grafica en un espacio geométrico. En la figura 34.2 se graficaron las cifras de la matriz factorial de la tabla 34.2.

Los dos factores, *A* y *B*, se colocan entre sí en un ángulo recto y se denominan *ejes de referencia*. Los valores apropiados de la carga del factor se señalan en cada uno de los ejes. Entonces cada una de las cargas de prueba se trata como una coordenada y se grafica. Por ejemplo, las cargas de la prueba *L* son (.79, .10). Se recorre hasta .79 sobre el eje *A* y se sube hasta .10 sobre el eje *B*. Este punto se indicó en la figura 34.2 por medio de una letra encerrada en un círculo, la cual indica la prueba. Las coordenadas de las otras cinco pruebas se grafican de manera similar.

FIGURA 34.2



La estructura factorial ahora se percibe con claridad. Cada prueba está altamente cargada en un factor, pero no en el otro. Todas son medidas relativamente "puras" de sus factores respectivos. Un séptimo punto se indicó en la figura 34.2 por medio de una *cruz* encerrada en un círculo, para ilustrar una supuesta prueba que mide ambos factores. Sus coordenadas son (.60, .50), lo cual significa que la prueba está cargada en ambos factores: .60 en *A* y .50 en *B*. No es "pura". Note que estructuras factoriales de esta simplicidad y claridad, donde 1) los factores son ortogonales (los ejes forman ángulos rectos entre sí), 2) las cargas de la prueba son sustanciales y "puras" (casi ninguna prueba se cargó con dos o más factores), y 3) sólo dos factores no son comunes. De nueva cuenta, el lector debe estar consciente de que el ejemplo no contiene datos reales.

La mayoría de los estudios de análisis de factores publicados reportan más de dos factores. Se han reportado cuatro, cinco e inclusive nueve, diez o más factores. La representación gráfica de dichas estructuras factoriales en una sola gráfica, por supuesto, no es posible. Por costumbre, los analistas factoriales grafican dos factores a la vez, aunque es posible graficar tres al mismo tiempo. No obstante, debe admitirse que es difícil visualizar o recordar estructuras *n*-dimensionales complejas. Por lo tanto, se visualizan estructuras bidimensionales y se generalizan a *n* dimensiones de forma algebraica. Un aspecto afortunado de los programas computacionales de análisis factorial es que dicha graficación de factores es fácilmente posible. Comrey (véase Comrey y Lee, 1992) desarrolló un programa gráfico para computadora que permite al usuario graficar dos factores al mismo tiempo.

Extracción y rotación de factores, puntuaciones factoriales y análisis factorial de segundo orden

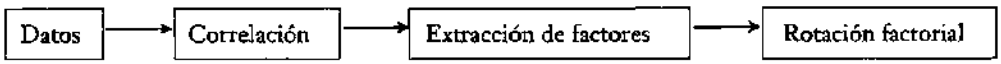
El análisis factorial moderno, tal y como lo define Thurstone (1947), implica cierto número de pasos. El primero consiste en la extracción de factores. Muchos de los métodos producen

factores que no son interpretables. Por consiguiente, dichos factores "sin rotación" se rotan con propósitos de interpretación. Existe un número de métodos de factores extraídos a partir de una matriz de correlación, que incluyen: factores principales, centroide, de probabilidad máxima, residual mínimo, de imagen, poder vectorial y alfa. No es posible analizar aquí todos estos métodos. El propósito de este texto es una comprensión básica y elemental, por lo que la explicación se limitará a uno de los métodos. El método más utilizado actualmente y que está fácilmente disponible en las instalaciones computacionales es el método de factores principales.

El lector puede preguntarse: ¿por qué no usar un método de grupo comparativamente simple, como el modelo de inspección utilizado con anterioridad, en lugar de un método complejo como el de factores principales? Los métodos de grupo pueden utilizarse (véase Lee y MacQueen, 1980) y se recomienda utilizarlos. Dependen de los grupos identificados y de los presuntos factores, por medio de encontrar grupos interrelacionados de coeficientes de correlación u otras medidas de relación. Resulta sencillo localizar los grupos en la tabla 34.1. Sin embargo, en la mayor parte de las matrices **R** no es tan fácil identificar los grupos. Se requieren métodos más objetivos y precisos.

En esta sección se examinan los principales pasos involucrados en el análisis factorial. No será posible presentar todos los detalles necesarios para hacerlo de forma completa. En su lugar, se espera brindarle al lector la esencia del análisis factorial y referirlo a muy buenos libros de texto para los detalles. Se presentarán buenos repasos de dichos textos en la sección de sugerencias de estudio.

Los pasos principales de los estudios que utilizan el análisis factorial se resumen de la siguiente forma:



El problema de la comunalidad y del número de factores

Antes de elegir qué método utilizar para la extracción de los factores, el investigador debe decidir qué va a poner en las casillas diagonales de la matriz de correlación como estimados comunales y cuántos factores va a extraer. Comrey (1978) señala que éstas son las dos decisiones más difíciles que un investigador debe tomar al hacer un análisis factorial. Lo que se utilice como estimados comunales y el número de factores a extraer, puede tener un fuerte impacto en la solución final (véase Comrey y Lee, 1992; Lee y Comrey, 1979; Lee, 1979). Si se conocen y utilizan las comunalidades correctas en el análisis factorial, se obtendrá el número correcto de factores, utilizando el método de factores principales que se describe en la siguiente sección. Por lo tanto, cuando un investigador usa el método de factores principales, los estimados comunales juegan un papel importante en la determinación de la solución factorial obtenida, y deben elegirse de manera cuidadosa. Los programas computacionales definitivamente han hecho que las soluciones factoriales por computadora sean más fáciles. Sin embargo, una de sus principales desventajas radica en que existen determinaciones dadas automáticamente por los programas para computadora, en términos de la forma en que se realiza la extracción factorial. Hubbard y Allen (1989) compararon las soluciones factoriales obtenidas por medio de dos populares programas computacionales, usando los valores predeterminados. Ellos encontraron soluciones muy diferentes. Algunos programas utilizan la regla del valor eigen, donde las unidades se ubican en la diagonal como estimados de las comunalidades y donde se extraen todos los factores que tienen un valor eigen igual o mayor que 1.00. Algunas veces dicho método se conoce como método de *componentes principales truncados*. Lo anterior tiene un

atractivo tanto intuitivo como matemático, ya que parece presentar una solución para ambos problemas. Sin embargo, Comrey y Lee (1992) han alertado en contra del uso indiscriminado de este método. Tiende a inflar demasiado las comunalidades y las cargas factoriales; entonces, las distorsiones se ven amplificadas por la rotación factorial. Aun así, continúa siendo uno de los procedimientos más utilizados. Las pruebas psicológicas y educativas que se desarrollaron gracias al uso de este método, como la *escala de estimación de habilidades sociales* (Social Skills Rating Scale) (Gresham y Elliot, 1990), deben interpretarse con precaución. Comrey y Lee (1992), así como Gorsuch (1983) han alertado sobre la realización del análisis factorial "a ciegas" y la posterior interpretación de los datos como verdaderos. Otro popular estimado de comunalidad es la correlación múltiple al cuadrado, R^2. La investigación realizada por Guttman (1956) demostró que dicho estadístico es el límite inferior para los estimados de las comunalidades y, como tal, podría subestimar los verdaderos valores de las comunalidades. Otros autores recomiendan el uso de la mayor correlación de la variable con otras variables, como el estimado inicial de la comunalidad.

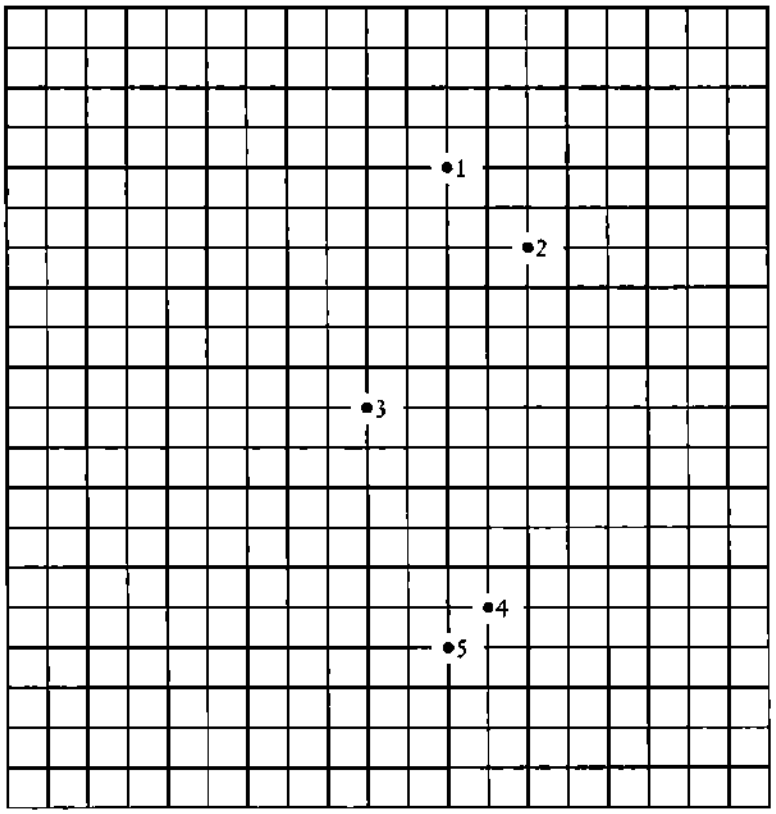
El método de factores principales

El método de factores principales es matemáticamente satisfactorio a causa de que produce una solución matemáticamente única de un problema factorial. Tal vez la principal característica de su solución es que extrae una cantidad máxima de varianza conforme se calcula cada factor. En otras palabras, el primer factor extrae la mayor cantidad de varianza, el segundo la siguiente mayor cantidad de varianza, y así sucesivamente. El primer factor consiste de pesos o coeficientes que maximizarán las correlaciones cuadradas entre las variables y el factor. Entonces la contribución del primer factor se remueve de la matriz de correlación. Entonces esta "nueva" matriz de correlación se usa para encontrar los coeficientes de un factor que maximice las correlaciones cuadradas entre las variables y el segundo factor. Cada factor subsecuente extraído tendrá cada vez menos varianza que el anterior a él. La extracción de factores cesa cuando la varianza se torna insignificante, o cuando el proceso de extracción alcanza el número de factores establecido por el investigador. Cada factor extraído consistirá en coeficientes que no están correlacionados con los coeficientes de los otros factores. En otras palabras, cada factor es independiente de los otros factores.

Es difícil demostrar la lógica del método de factores principales sin demasiadas matemáticas. No obstante, se puede lograr una cierta comprensión intuitiva del método al enfocarlo de manera geométrica. Conciba las pruebas o variables como puntos en un espacio m-dimensional. Las variables que están alta y positivamente correlacionadas deben estar cercanas entre sí, y lejos de las variables con las que no se correlacionan. Si tal razonamiento es correcto, debe haber conjuntos de puntos en el espacio. Cada uno de esos puntos puede ser ubicado en el espacio si se insertan ejes adecuados en éste, es decir, un eje para cada dimensión de las m dimensiones. Entonces, cualquier ubicación de un punto es su identificación múltiple obtenida por medio de la lectura de sus coordenadas en los ejes m. El problema factorial consiste en proyectar los ejes a través de conjuntos vecinos de puntos, para así ubicar aquellos ejes que "explican" tanta varianza de las variables como sea posible.

Es posible demostrar lo anterior con un ejemplo bidimensional simple. Suponga que se tienen cinco pruebas y que están situadas en un espacio bidimensional, tal como se indica en la figura 34.3. Cuanto más cerca estén dos puntos, tendrán mayor relación. El problema es determinar: 1) cuántos factores hay, 2) cuáles pruebas están cargadas en cuáles factores y 3) la magnitud de las cargas de las pruebas.

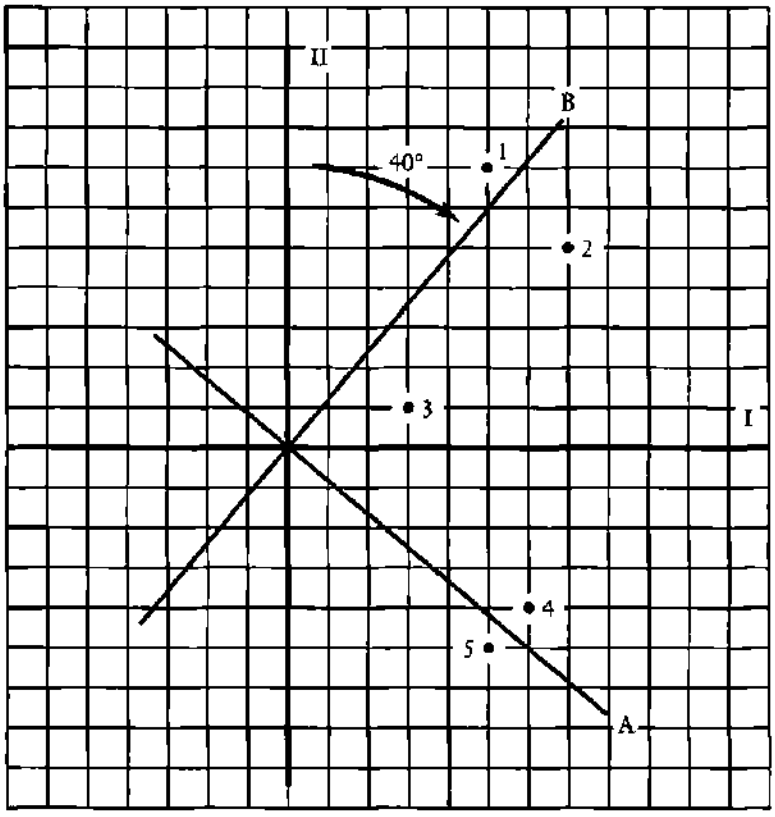
FIGURA 34.3



Ahora el problema se va a resolver de dos maneras diferentes, cada una interesante e instructiva. Primero, se resuelve directamente a partir de los propios puntos. Es menester seguir las siguientes instrucciones. Trace una línea vertical tres unidades a la izquierda del punto 3. Dibuje una línea horizontal debajo del punto 3. Designe estos ejes de referencia I y II. Ahora lea las coordenadas de cada punto, por ejemplo, el punto 2 es (.70, .50), el punto 4 es (.60, –.40). Escriba una "matriz factorial" con estos cinco pares de valores.

Después los ejes se rotan ortogonalmente y en sentido de las manecillas del reloj, de tal manera que el eje I quede entre los puntos 4 y 5. En efecto, el eje II queda entre los puntos 1 y 2. (Se recomienda el uso de un transportador; la rotación debe ser de aproximadamente 40°.) A estos "nuevos" ejes rotados se les llama *A* y *B*. Ahora se corta una tira de papel de cuadrícula pequeña. (Los puntos se grafican en papel cuadriculado.) Considere la base de cada cuadrado como .10 (.10 = ¼ de pulgada; 10 unidades de hecho son iguales a 1.00). Utilizando la tira como instrumento de medición, se miden las distancias de los puntos con los nuevos ejes. Por ejemplo, el punto 2 debe estar cerca de (.22, .83), y el punto 5 debe estar cerca de (.71, –.06). (No importa si existen pequeñas discrepancias.) Los ejes originales (I y II) y los rotados (*A* y *B*) y los cinco puntos se presentan en la figura 34.4.

FIGURA 34.4



Ahora se escriben ambas matrices factoriales, sin rotar y rotadas. Éstas se presentan en la tabla 34.4. El problema está resuelto: hay dos factores. Los puntos (pruebas) 1 y 2 tienen cargas altas del factor *B*, los puntos 4 y 5 tienen cargas altas del factor *A*, y el punto 3 tiene cargas bajas de ambos factores. Se respondieron las tres preguntas planteadas originalmente.

TABLA 34.4 *Matrices sin rotación y rotada, problema de la distancia de los puntos*[a]

Puntos	Sin rotación		Puntos	Rotada	
	I	II		A	B
1	.50	.70	1	−.07	*.86*
2	.70	.50	2	.22	*.83*
3	.30	.10	3	.17	.27
4	.60	−.40	4	*.72*	.08
5	.50	−.50	5	*.71*	−.06

[a] Las cargas con rotación sustanciales están en *itálicas*.

TABLA 34.5 *Matrices factoriales sin rotación y rotadas, la matriz R de la tabla 36.1*[a]

	Sin rotación		Rotadas		
Pruebas	I	II	*A*	*B*	h^2
V	.60	−.58	*.83*	.01	.70
L	.63	−.49	*.79*	.10	.63
S	.56	−.43	*.70*	.10	.50
N	.56	.43	.10	*.70*	.50
AE	.63	.49	.10	*.79*	.63
AEP	.60	.58	.01	*.83*	.70

[a] Las cargas significativas > 2.30 están en *itálicas*.

Este procedimiento es análogo a los problemas factoriales psicológicos. Las pruebas se consideran como puntos en un espacio factorial *m*-dimensional. Las cargas factoriales son las coordenadas. El problema es introducir marcos de referencia o ejes apropiados y después "leer" las cargas factoriales. Por desgracia, en problemas reales no se conoce el número de factores (la dimensionalidad del espacio factorial y, por lo tanto, el número de ejes) o la ubicación de los puntos en el espacio, los cuales es necesario determinar a partir de los datos.

La descripción anterior es figurativa. No se "leen" las cargas factoriales a partir de los ejes de referencia; se calculan utilizando métodos bastante complejos. El método de factores principales en realidad involucra la solución de ecuaciones lineales simultáneas. Las raíces obtenidas de las soluciones se denominan *valores eigen*. También se obtienen *vectores eigen*; después de la transformación adecuada, se convierten en cargas factoriales. Así se resolvió la matriz **R** ficticia de la tabla 34.1 y produjo la matriz factorial que se presentó después en la tabla 34.5. La mayoría de los programas de análisis computacionales usan las soluciones de factores principales. El estudiante que tenga la expectativa de utilizar el análisis factorial con cualquier profundidad, debe estudiar el método cuidadosamente y, por lo menos, entender lo que hace. No hay nada tan riesgoso y autoderrotante como usar ciegamente programas computacionales. En especial, éste es el caso en el análisis factorial.

Rotación y estructura simple

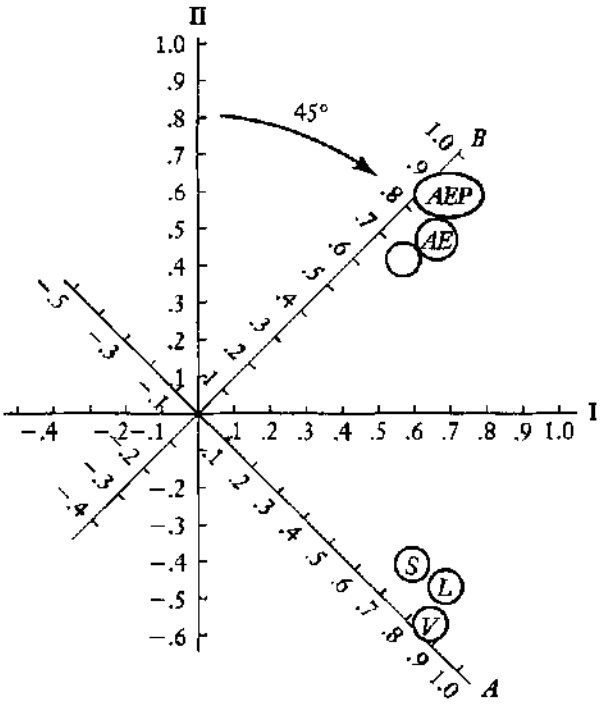
La mayoría de los métodos de extracción factorial producen resultados de tal forma que son difíciles o imposibles de interpretar. Lo anterior se percibe al observar los factores sin rotación de la tabla 34.4. Thurstone (1947, pp. 508-509) comentó que era necesario rotar las matrices factoriales si se deseaba interpretarlas de manera adecuada. Señaló que las matrices factoriales originales son arbitrarias en el sentido de que es posible encontrar un infinito número de marcos de referencia (ejes), para reproducir cualquier matriz **R** dada (véase Thurstone, 1947, p. 93). Una matriz de factores principales y sus cargas explican la varianza del factor común de las puntuaciones de la prueba; pero en general no proporcionan estructuras con un significado científico. Son las configuraciones de las pruebas o variables en el espacio factorial las que tienen una importancia fundamental. Para descubrir tales configuraciones de manera adecuada, deben rotarse los ejes de referencia arbitrarios. En otras palabras, se supone que existen posiciones únicas y "mejores" de los ejes, "mejores" formas de ver las variables en el espacio *n*-dimensional.

No existe aquí la intención de materializar constructos, variables o factores. Los factores son meras estructuras o patrones generados por covarianzas de mediciones. Lo que se quiere decir con "mejores maneras de ver las variables" es la manera más parsimoniosa y más simple. Una "mejor" forma se predice a partir de la teoría y de las hipótesis. O una "mejor" forma se descubre a partir de una estructura tan clara y fuerte que casi obligue a creer en su validez y "realidad".

Entre las importantes contribuciones de Thurstone, la invención de las ideas de la estructura simple y de la rotación de los ejes factoriales son, tal vez, las más importantes. Con ellas, él estableció lineamientos relativamente claros para lograr soluciones analíticas factoriales con significado psicológico e interpretables. En la tabla 34.2 se reportó una matriz factorial obtenida a partir de la matriz **R** de la tabla 34.1. Ésta era la matriz ***rotada*** **final y no la matriz producida originalmente por el análisis factorial.** La matriz ***sin rotación***, originalmente producida por medio del método de factores principales, se presenta en la parte izquierda de la tabla 34.5. Los factores rotados se reproducen en la parte derecha de la tabla. También se muestran las comunalidades (h^2) que son iguales en las dos matrices.

Si se intenta interpretar la matriz sin rotación en la parte izquierda de la tabla, se enfrenta un problema. Se puede decir que todas las pruebas se cargan de forma sustancial sobre un factor I general; y que el segundo factor, II, es bipolar. (Un *factor bipolar* es aquel que tiene cargas sustanciales positivas y negativas.) Lo anterior equivale a decir que todas las pruebas miden lo mismo (factor I), pero que las tres primeras miden el aspecto negativo de lo que sea que miden las otras tres (factor II). Pero aparte de la naturaleza ambigua

FIGURA 34.5



de una interpreración como ésta, se sabe que los ejes de referencia, I y II, y en consecuencia las cargas factoriales, son arbitrarios. Observe la gráfica factorial de la figura 34.2. Existen dos grupos de pruebas claramente definidos, pegados cerca de los ejes *A* y *B*. Aquí *no* hay un factor general, tampoco hay un factor bipolar. El segundo problema principal del análisis factorial, por lo tanto, consiste en descubrir una solución única y convincente o la posición de los ejes de referencia.

Si se grafican las cargas de I y II se "observa" la estructura original sin rotación. Esto se hizo en la figura 34.5. Ahora se giran los ejes de tal manera que I queda lo más cerca posible de los puntos *V, L* y *S*, al mismo tiempo, II queda lo más cerca posible a los puntos *N, AE* y *AEP*. Una rotación de 45° será adecuada. Entonces se obtiene, en esencia, la estructura de la figura 34.2. Es decir, las nuevas posiciones con los ejes rotados y las posiciones de las seis pruebas son iguales a las posiciones de los ejes y pruebas de la figura 34.2. La estructura simplemente se inclina a la derecha. Si se gira la figura, de tal manera que la *B* del eje *B* apunte directamente hacia arriba, esto se hace evidente. Ahora es posible leer las nuevas cargas factoriales rotadas sobre los ejes rotados. Puesto que los ejes se mantienen en un ángulo recto de 90°, se denomina rotación *ortogonal*.

Este ejemplo, aunque poco realista, puede ayudar al lector a comprender que los analistas factoriales buscan las unidades que presuntamente subyacen al desempeño de las pruebas. Concebido de forma espacial, buscan las relaciones entre variables "afuera" en el espacio factorial multidimensional. Por medio del conocimiento de las relaciones empíricas entre pruebas u otras medidas, exploran el espacio factorial con los ejes de referencia hasta que encuentran las unidades o relaciones entre relaciones —si es que existen—.

Las cargas significativas ($\geq .30$) están en itálicas. Note que los vectores *A* y *B* están invertidos en esta tabla. Las h^2 calculadas a partir de los valores sin rotación y con rotación difieren ligeramente, a causa de los errores de redondeo; por ejemplo, $.60^2 + .58^2 = .70$, y $.83^2 + .01^2 = .69$. En la tabla se utilizaron los valores exactos obtenidos por computadora (y en la tabla 34.2).

Para dirigir las rotaciones, Thurstone estableció cinco principios o reglas de la estructura simple. Las reglas son aplicables tanto para las rotaciones ortogonales como para las oblicuas; aunque Thurstone enfatizó el caso oblicuo. (Las rotaciones oblicuas son aquellas donde los ángulos entre los ejes son agudos y obtusos.) Los principios de la estructura simple son los siguientes:

1. Cada renglón de la matriz factorial debe tener por lo menos una carga cercana a cero.
2. Por cada columna de la matriz factorial debe haber por lo menos tantas variables con cargas iguales o cercanas a cero como factores.
3. Por cada par de factores (columnas) debe haber diversas variables con cargas en un factor (columna), pero no en el otro.
4. Cuando haya cuatro o más factores, una gran proporción de las variables debe tener cargas insignificantes (cercanas a cero) en cualquier par de factores.
5. Por cada par de factores (columnas) de la matriz factorial debe haber sólo un pequeño número de variables con cargas sustanciales (diferentes de cero), en ambas columnas.

En efecto, dichos criterios demandan variables que sean lo más "puras" posibles; es decir, que cada variable esté cargada en el menor número de factores posibles y que haya *la mayor cantidad de ceros posibles en la matriz factorial rotada*. De esta forma, es posible lograr la interpretación más simple posible de los factores. En otras palabras, la rotación para lograr la estructura más simple es una manera bastante objetiva de lograr la simplicidad de variables o reducir la complejidad de las variables.

Para comprender lo antes expuesto, imagine una solución ideal en donde la estructura simple sea "perfecta". Podría verse de la siguiente forma, por ejemplo, en una solución de tres factores, la cual se presenta en la siguiente página.

Las X indican cargas factoriales sustanciales, los 0 indican cargas cercanas a cero. Por supuesto, dichas estructuras factoriales "perfectas" son poco frecuentes. Es más probable que algunas de las pruebas tengan cargas en más de un factor. Aun así, se han logrado buenas aproximaciones a la estructura simple, especialmente en estudios analíticos factoriales bien planeados y bien ejecutados. Comrey (1978) señala que la estructura simple funciona bien si el estudio está bien diseñado, con varios factores bien definidos y donde cada uno se mide a través de varias medidas factoriales puras que están distribuidas de manera normal y con alta confiabilidad. Sin embargo, Comrey también afirma que estudios sin diseño o mal diseñados tendrán variables complejas; y como resultado, la solución no se ajustará muy bien a la estructura simple.

Antes de abandonar el tema de la rotación factorial, debe señalarse que existe una serie de métodos de rotación. Los dos tipos principales de rotación son los llamados "ortogonal" y "oblicuo". Las rotaciones *ortogonales* mantienen la independencia de los factores; es decir, los ángulos entre los ejes se mantienen a 90°. Por ejemplo, si se rotan los factores I y II de forma ortogonal, se giran juntos ambos ejes, manteniendo el ángulo recto entre ellos. Esto quiere decir que la correlación entre los factores es cero. La rotación realizada en la figura 34.5 es ortogonal. Si se tuvieran cuatro factores, se rotarían I y II, I y III, I y IV, II y III, etcétera, manteniendo ángulos rectos entre cada par de ejes. Algunos investigadores prefieren hacer una rotación ortogonal. Otros insisten en que la rotación ortogonal no es realista, que los factores reales por lo general están correlacionados, y que las rotaciones deben ajustarse a la "realidad" psicológica.

Las rotaciones donde los ejes factoriales permiten formar ángulos agudos u obtusos se denominan *oblicuas*. Por supuesto, oblicuo significa que los factores están correlacionados. No cabe duda de que las estructuras factoriales pueden ajustarse mejor con ejes oblicuos y que los criterios de la estructura simple se cumplen mejor. Algunos investigadores objetan los factores oblicuos debido a la posible dificultad al comparar estructuras factoriales de un estudio a otro. Se deja el polémico tema con dos señalamientos. Primero, el tipo de rotación parece ser una cuestión de gusto. Segundo, el lector necesita comprender ambos tipos de rotación al grado de que pueda interpretar ambos tipos de factores, y ser particularmente cuidadoso al enfrentarse con los resultados de soluciones oblicuas, los cuales contienen particularidades y sutilezas que no están presentes en las soluciones ortogonales.

La rotación factorial que se ha estudiado hasta ahora constituye el modelo gráfico. Antes de la existencia de las computadoras digitales de alta velocidad, los investigadores que realizaban análisis factoriales usaban dicho método gráfico o manual de rotación. La naturaleza imprecisa de las rotaciones gráficas por medio de aproximaciones visuales, era una de las principales críticas al método. Sin embargo, conforme las computadoras se volvieron más precisas y confiables a la mitad de los años cincuenta, emergió un número de métodos analíticos de rotación, donde las rotaciones se realizaban utilizando una fórmula matemática. El más popular fue el desarrollado por Henry Kaiser (1958) llamado Varimax. Virtualmente cada paquete computacional que lleva a cabo análisis factoriales en la actualidad, utiliza este método de rotación. En muchos de tales programas el método Varimax se utiliza de forma automática. En la esquela de defunción de Kaiser, Jensen y Wilson (1994) afirmaron que el artículo de Kaiser de 1958 es el tercer artículo más citado en la literatura psicológica. Varimax funciona muy bien en la aproximación de la estructura simple en estudios analíticos factoriales bien diseñados. Si un investigador está buscando un factor general, el método Varimax no funciona muy bien. La meta del método Varimax consiste en dispersar la mayor cantidad de varianza a través de los factores y, al mismo

Pruebas	A	B	C
1	X	0	0
2	X	0	0
3	X	0	0
4	0	X	0
5	0	X	0
6	0	X	0
7	0	0	X
8	0	0	X
9	0	0	X

tiempo, tratar de obtener la estructura simple. Si el investigador incluye demasiados factores al emplear el método Varimax, entonces un resultado posible es el incremento artificial de los factores pequeños. Por lo tanto, Varimax se vuelve sensible al número de factores utilizados en la rotación.

Si un investigador está interesado en encontrar un factor general, el mejor método ortogonal disponible es el *criterio de tándem de Comrey* (1967), que consiste en dos pasos, cada uno de los cuales se basa en un principio diferente.

> Principio 1: Si dos variables están correlacionadas, deben aparecer en el mismo factor (criterio 1).
> Principio 2: Si dos variables no están correlacionadas, no deben aparecer en el mismo factor (criterio 2).

El *criterio* o *principio 1* es el más interesante si se busca un factor general. El *criterio 1* intenta dispersar la varianza desde los factores más grandes hasta los más pequeños, mientras satisface el *principio 1* al mismo tiempo. Si existe un factor general, las variables que estén correlacionadas entre sí, serán retenidas lo más posible en el mismo factor, en lugar de dispersarse alrededor.

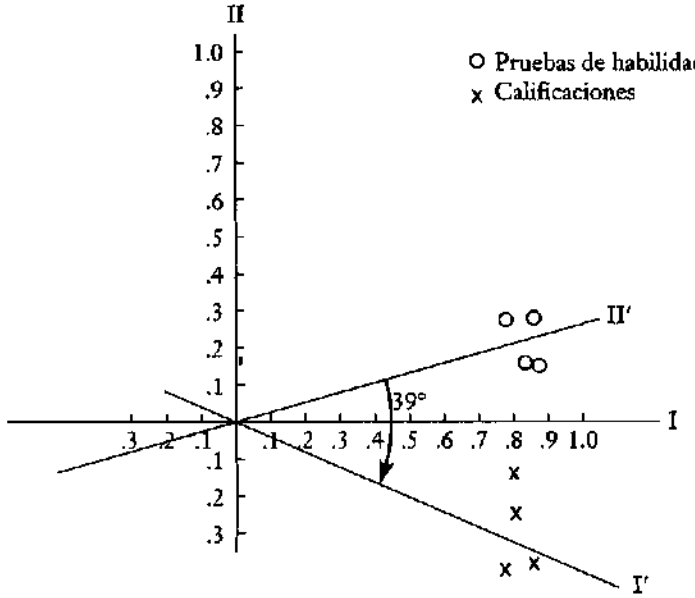
Del mismo modo en que existen muchos métodos de extracción de factores, también hay muchos métodos de rotación, además de los que ya se han mencionado. Existen diferentes métodos tanto para la rotación ortogonal como para la oblicua. El lector puede consultar a Gorsuch (1983) y a Comrey y Lee (1992), donde encontrará una lista.

Análisis factorial de segundo orden

El análisis factorial de segundo orden es un método sumamente importante, aunque rechazado, para el análisis de datos complejos y la comprobación de hipótesis. Cuando se rotan los datos de forma oblicua, existen correlaciones entre los factores. Antes en el presente capítulo se mencionó la ecuación fundamental del análisis factorial. Una versión de ella era $\mathbf{R} = \mathbf{P}\,\Phi\,\mathbf{P}^T + \mathbf{U}$. La matriz Φ contiene la correlación entre factores. En las rotaciones ortogonales, la matriz Φ no se utiliza debido a que los factores no están correlacionados. Para realizar un análisis factorial de segundo orden de la manera tradicional, dicha matriz Φ se analiza factorialmente. Por completar, la matriz $\mathbf{P}$ es la matriz del patrón factorial; contiene las cargas factoriales. $\mathbf{P}^T$ es su transposición y $\mathbf{U}$ es la matriz que contiene la singularidad de cada variable.

En un estudio provocativo de análisis factorial y de correlación canónica, sobre la redundancia presente en puntuaciones de pruebas de estudiantes, Lohnes y Marshall (1965)

FIGURA 34.6



extrajeron dos factores de 21 pruebas de habilidad y rendimiento. Las cargas sin rotación de ocho de sus medidas, cuatro pruebas de habilidad y cuatro calificaciones (inglés, aritmética, estudios sociales, ciencia) se graficaron en la figura 34.6. Los ejes se rotaron de forma oblicua, de tal manera que quedaran entre los dos grupos de cargas. Existe un ángulo agudo de aproximadamente 39° entre los ejes rotados, ahora denominados I´ y II´. Cualquier ángulo entre los ejes que no sea de 90° implica la existencia de correlación entre los factores. En este caso, la correlación es bastante alta, aproximadamente de .78.

Imagine la situación anterior multiplicada por seis, ocho o 10 factores: habría un conjunto de correlaciones entre los factores. Si se analizan factorialmente estas correlaciones se tiene un análisis factorial de segundo orden, que es un método para encontrar los factores que subyacen a los factores. El famoso componente *g* de las pruebas de inteligencia es, evidentemente, un factor de segundo orden o de orden mayor. Siempre que se analizan factorialmente grandes cantidades de pruebas de habilidad, las correlaciones entre las pruebas son, por lo general, positivas. Si se analizan factorialmente surge un patrón como éste, aunque más complejo, tal como sucedió en la figura 34.6. Si se calculan las correlaciones entre los factores y se analizan factorialmente de nuevo, puede surgir un solo factor, quizás *g*.

Para obtener información adicional sobre el análisis factorial de segundo orden y de orden mayor véase Gorsuch (1983). El investigador puede realizar el proceso para encontrar factores oblicuos de primer orden y después analizar factorialmente la matriz de correlación de factores; aunque existe un método alternativo. Con la disponibilidad de programas para computadora, tales como LISREL, EQS y AMOS, el investigador puede realizar un análisis factorial de orden superior de forma bastante fácil y en un solo paso. Comrey y Lee (1992) demuestran cómo realizar un análisis factorial de segundo orden utilizando el programa EQS.

Puntuaciones factoriales

Mientras que el análisis factorial de segundo orden está más orientado hacia la investigación básica y teórica, otra técnica de análisis factorial, las llamadas puntuaciones o medidas factoriales, es eminentemente práctica, pero no sin importancia teórica. Las *puntuaciones factoriales* son medidas de los individuos en los factores. Suponga que, como lo hicieron Lohnes y Marshall (1965), se encuentran dos factores detrás de 21 medidas de habilidad y de calificación. En lugar de utilizar las 21 puntuaciones de los grupos de niños en investigación, ¿por qué no utilizar sólo dos puntuaciones calculadas a partir de los factores? Lohnes y Marshall recomiendan hacerlo así, y señalan la redundancia en las puntuaciones usuales de los alumnos. En efecto, dichas puntuaciones factoriales son promedios ponderados: ponderados de acuerdo a las cargas factoriales.

A continuación se presenta un ejemplo sobresimplificado. Suponga que los datos de la matriz factorial de la tabla 34.2 fueran reales y que se desea calcular las puntuaciones factoriales *A* y *B* de un individuo. Las puntuaciones en bruto de un individuo en las seis pruebas son, por ejemplo: 7, 5, 5, 3, 4 y 2. Se multiplican tales puntuaciones por las cargas factoriales relacionadas, primero para el factor *A* y después para el factor *B*, de la siguiente manera:

$$A\text{: } FA = (.83)(7) + (.79)(5) + (.70)(5) + (.10)(3) + (.10)(4) + (.01)(2) = 13.98$$

$$B\text{: } FB = (.01)(7) + (.10)(5) + (.10)(5) + (.70)(3) + (.79)(4) + (.83)(2) = 7.99$$

Las "puntuaciones factoriales" del individuo son $FA = 13.98$ y $FB = 7.99$. Por supuesto, es posible calcular las "puntuaciones factoriales" de otros individuos de manera similar.

Ésta no es la mejor forma para calcular las puntuaciones factoriales. Comrey y Lee (1992), Gorsuch (1983) y Harman (1976) presentan métodos alternativos para calcular las puntuaciones factoriales. Ellos también explican las ventajas y desventajas de cada método. Sin embargo, el ejemplo aquí presentado fue inventado solamente para trasmitir la idea de dichas puntuaciones como sumas o promedios ponderados, donde los pesos son las cargas factoriales. En cualquier caso, aunque el método no fue utilizado de forma extensa en el pasado, tiene un gran potencial para la investigación compleja del comportamiento. En lugar de utilizar muchas puntuaciones separadas, se utiliza un menor número de puntuaciones factoriales. Un excelente ejemplo real es el descrito por Mayeske (1970), quien participó en un nuevo análisis de los datos del reporte de gran influencia de Coleman, Campbell, Hobson, *et al.* (1966) llamado *Equality of Educational Opportunity.*

Ejemplos de investigación

La mayoría de los estudios analíticos factoriales han factorizado las pruebas y escalas de inteligencia, aptitud y personalidad, donde se han intercorrelacionado y analizado factorialmente las propias pruebas y escalas. El ejemplo de Thurstone que se explicó anteriormente es un excelente ejemplo; de hecho, es un clásico. También se pueden factorizar las personas o sus respuestas. De hecho, las variables incluidas en las matrices de correlación y factoriales pueden ser pruebas, escalas, personas, reactivos, conceptos o cualquier cuestión que pueda estar intercorrelacionada. Los estudios descritos más adelante han sido seleccionados no para representar investigaciones analíticas factoriales en general, sino más bien para familiarizar al estudiante con los diferentes usos del análisis factorial.

Las escalas de personalidad de Comrey

El trabajo de Comrey (1970) sobre investigación de la personalidad constituye uno de los mejores ejemplos sobre el uso del análisis factorial. Las escalas de personalidad de Comrey, también conocidas como el CPS (por sus siglas en inglés), conforman un inventario de rasgos de personalidad factorizados. Dicha taxonomía de rasgos de personalidad se desarrolló durante un periodo de 15 años. Originalmente se inspiró en las discrepancias existentes entre reconocidos autores de pruebas de personalidad y teóricos de la personalidad. Lo que inició como un esfuerzo para resolver las diferencias entre los teóricos de la personalidad finalizó con el surgimiento de las escalas de personalidad de Comrey. Las CPS comparten algunas de las características de las otras pruebas; pero difieren de ellas.

Para obtener una solución factorial estable y control de la jerarquía factorial, Comrey desarrolló una unidad de medición llamada "dimensión del reactivo homogéneo factorizado" (DRHF), que se desarrolló para resolver algunos de los problemas experimentados en los reactivos factorizados. Un problema se refiere a la falta de confiabilidad asociada con reactivos únicos. La otra se refiere a la extracción de factores de reactivos que sean factores de bajo nivel, consistentes de reactivos que sean similares en su redacción u otra características distinguible. Tales factores de reactivos de bajo nivel, por lo general, no ofrecen al investigador mucha información sobre el rasgo de personalidad subyacente. La DRHF no es más que la suma de las puntuaciones de los reactivos que definen la DRHF. Puesto que la DRHF es una suma, es más confiable que cualquier reactivo solo.

Un análisis factorial de la DRHF produce ocho factores. Los nombres de estos factores son *confianza* contra *defensividad*; *disciplina* contra *falta de compulsión*; *conformidad social* contra *rebelión*; *actividad* contra *falta de energía*; *estabilidad emocional* contra *neurosis*; *extraversión* contra *introversión*; *fuerza mental* contra *sensibilidad*; y *empatía* contra *egocentrismo*.

Información adicional sobre los pasos y procedimientos del desarrollo de las CPS puede encontrarse en Comrey y Lee (1992), Comrey (1980) y Comrey (1988). De los años setenta a los años noventa Comrey y sus colegas han validado dicha estructura de ocho factores en varias culturas diferentes y países diferentes. La investigación actual realizada por Comrey sobre las CPS demuestra todos los pasos correctos que toma un investigador al realizar un estudio analítico factorial.

Estudio factorial de Thurstone sobre la inteligencia

Thurstone y Thurstone (1941), en su trabajo monumental sobre los factores de inteligencia y su medición, analizaron factorialmente 60 pruebas además de las variables *edad cronológica*, *edad mental* y *sexo*. El análisis se basó en las respuestas de 710 alumnos de primer año de secundaria a 60 pruebas y reveló, esencialmente, el mismo conjunto de los llamados factores primarios, que se habían encontrado en estudios previos.

Los Thurstone eligieron las tres mejores pruebas de cada uno de siete de los 10 factores primarios. Seis de estas pruebas parecían tener estabilidad a distintos niveles de edad, para usos escolares prácticos. Entonces, ellos revisaron y administraron las pruebas a 437 estudiantes de segundo grado de secundaria. El principal propósito del estudio consistía en verificar la estructura factorial de las pruebas. En otras palabras, ellos predijeron que los mismos factores primarios de inteligencia puestos en las 21 pruebas surgirían de un nuevo análisis factorial, en una nueva muestra de niños.

Inteligencia fluida y cristalizada

Uno de los problemas más activos, importantes y polémicos de interés científico y práctico del comportamiento es la naturaleza de las habilidades mentales. Diferentes teorías con diversas cantidades de tipos de evidencia que las soporten han sido propuestas por algunos de los psicólogos más sobresalientes del siglo: Spearman, Thurstone, Burt, Thorndike,

Guilford, Cattell y otros. No cabe ninguna duda respecto a la gran importancia científica y práctica del problema. Se ha aludido, aunque sólo brevemente, al trabajo y pensamiento de Thurstone y Guilford. Ahora se describirá, también de manera breve, uno de los muchos estudios analíticos factoriales de Cattell (1963).

Puede mostrarse que el famoso factor general de inteligencia, *g*, es un factor de segundo orden que aparece en la mayoría de las pruebas de habilidad mental. En efecto, Cattell considera que existen dos *g*, o dos aspectos de *g*: el cristalizado y el fluido. La *inteligencia cristalizada* se exhibe por medio de desempeños cognitivos donde "hábitos de juicio hábiles" se han fijado o cristalizado, debido a la aplicación anterior de la habilidad de aprendizaje general a dichos desempeños. Los reconocidos factores verbal y numérico son ejemplos de ello. Por otro lado, la *inteligencia fluida* se exhibe por medio de desempeños caracterizados más por la adaptación a situaciones nuevas, la aplicación "fluida" de la habilidad general, por así decirlo. Dicha habilidad es más característica del comportamiento creativo que la inteligencia cristalizada. Si los factores se analizan factorialmente y las correlaciones entre factores que se encuentren se factorizan a su vez (análisis factorial de segundo orden), entonces tanto la inteligencia cristalizada como la fluida deben surgir como factores de segundo orden.

Cattell administró la prueba de habilidades primarias de Thurstone y un número de sus propias pruebas de habilidad mental y personalidad a 277 niños de segundo grado de secundaria, después analizó factorialmente las 44 variables y rotó los 22 factores obtenidos (probablemente demasiados) a la estructura simple. Las correlaciones entre dichos factores fueron, a su vez, factorizadas, produciendo ocho factores de segundo orden. (Recuerde que las rotaciones oblicuas producen factores que están correlacionados.) A pesar de que Cattell incluyó un número de variables de personalidad, aquí se enfocan sólo los primeros dos factores: inteligencia fluida e inteligencia cristalizada. Él pensó que las pruebas de Thurstone debían cargarse en un factor general, puesto que miden habilidades cognitivas cristalizadas; y que sus propias pruebas relacionadas con la cultura debían cargarse en otro factor, debido a que miden habilidad fluida. Y así sucedió. Los dos conjuntos de cargas factoriales se indican en la tabla 34.6, junto con los nombres de las pruebas. Los dos factores estuvieron también correlacionados positivamente ($r = .47$), como se predijo.

Dicho estudio demuestra el poder de una inteligente combinación de teoría, construcción de pruebas y análisis factorial. Similar a la también inteligente conceptualización

TABLA 34.6 *Parte de la matriz factorial de segundo orden (estudio de la inteligencia fluida y cristalizada de Cattell)*[a]

	$F_1(gf)$	$F_2(gc)$
Pruebas de Thurstone:		
Verbal	.15	*.46*
Espacial	*.32*	.14
Razonamiento	.08	*.50*
Numérica	.05	*.59*
Fluidez	.07	.09
Pruebas de Cattell:		
Series	*.35*	*.43*
Clasificación	*.63*	−.02
Matrices	*.50*	.10
Topología	*.51*	.09

[a] *gf* = factor general fluido; *gc* = factor general cristalizado. Las itálicas fueron añadidas por el autor (FNK). Éstos son sólo dos de los ocho factores de Cattell.

y análisis de factores divergentes, convergentes y de otros tipos ya mencionados, realizada por Guilford, se trata de una significativa contribución al conocimiento psicológico de un tema extremadamente complejo e importante. Sin embargo, para obtener una visión completa, también debe leerse el artículo de Humphreys (1967) que critica la teoría de Cattell.

Análisis factorial confirmatorio

Los modelos de análisis factorial descritos anteriormente hasta este punto representan procedimientos tradicionales que, por lo general, ahora se conocen como "análisis factorial exploratorio" o AFE. Se han creado métodos más novedosos con una base más fuerte sobre la teoría de comprobación de hipótesis, los cuales se conocen como "análisis factorial confirmatorio". Sin embargo, existen pequeñas variantes del análisis factorial confirmatorio. Existen los iniciales, basados en el AFE y los más nuevos basados en teoría estadística más estricta. Los métodos más nuevos serán referidos como AFC.

Anteriormente en este capítulo se enfatizó que los métodos analíticos factoriales exploratorios son muy poderosos cuando se utilizan para la comprobación de hipótesis. Es decir, cuando se desarrollan hipótesis tanto acerca de los factores que se busca encontrar en cierto dominio como acerca de las variables que los miden. Se eligen diversas variables para cada factor hipotetizado que debe proporcionar medidas relativamente puras de ese factor. Se reúnen datos para una muestra grande y se analizan factorialmente para ver qué tan bien los factores obtenidos y las variables cargadas en ellos corresponden a la estructura factorial originalmente hipotetizada. Con base en el primer análisis, se efectúan revisiones en la hipótesis y en las variables designadas para medir cada factor, y se repite el estudio. Dicho proceso se repite de manera pragmática hasta que la estructura factorial que emerge corresponde razonablemente bien a la estructura factorial hipotetizada con anticipación.

Se trata de un método que representa el tipo anterior de análisis factorial confirmatorio donde el número de factores que surgen a partir del análisis no está restringido a un número preconcebido. Si el número "correcto" resulta ser el que se hipotetizó, eso está bien, pero no se especifica con antelación. Además, se permite que las cargas caigan donde sea, en lugar de ser forzadas a conformar lo más posible un patrón especificado previamente. En particular, grandes números de parámetros no se fuerzan hacia cero. La proporción de la varianza atribuida a factores únicos para cada variable surge como un resultado final del análisis, en lugar de ser un parámetro, *per se*, a estimarse. Así, en el AFE se evalúa qué tan bien se ajusta la solución obtenida a un patrón factorial preconcebido; aunque sin el uso de poderosas técnicas de optimización que fuercen un ajuste que puede tener un débil sostén con datos nuevos.

Algunos de esos métodos iniciales AFE del análisis factorial confirmatorio se denominan soluciones *de rotación forzada o soluciones procusteanas** (véase Comrey y Lee, 1992). Dicho método es susceptible de adoptar diversas formas. Se puede rotar una matriz no rotada de la manera más cercana posible a una matriz meta, ya sea ortogonal u oblicua. La matriz meta podría ser una matriz hipotética basada en la teoría o en las expectativas desarrolladas a partir de investigación previa. Por lo general, se utilizan métodos de mínimos cuadrados para encontrar la matriz de transformación que logrará las rotaciones de-

*Nota del revisor técnico: Como lo mencionan Nunnally y Bernstein en su libro *Teoría psicométrica* de esta misma editorial: "El nombre Procusto viene de un posadero de la mitología griega que tenía una cama que se ajustaría al tamaño de cualquiera. Si el visitante era demasiado pequeño para la cama, Procusto alargaba al visitante en un potro. Si un visitante era demasiado alto para caber en la cama, Procusto acortaba la longitud de las piernas del visitante para que se adaptara a la cama" (pág. 630).

seadas. El estudio de Verba y Nie (1972) constituye un ejemplo de un análisis factorial confirmatorio realizado con el uso de tal método.

Esos antiguos métodos de análisis factorial confirmatorio tal vez se volverán menos populares, a medida que los métodos más novedosos sean capaces de realizar el mismo tipo de tarea en la mayoría de los casos y, además, proporcionen una prueba estadística de bondad de ajuste, así como indicaciones sobre cómo mejorar el modelo.

Los métodos más nuevos, que aquí se denominan AFC, se basan en el trabajo de Lawley (1940), quien introdujo el método de probabilidad máxima al análisis factorial y posteriormente lo desarrolló (véase Lawley y Maxwell, 1971). Gorsuch (1983) señala que el AFC tiene sus raíces en el método de análisis factorial de probabilidad máxima de Maxwell-Lawley. Dichos métodos nuevos por lo común destacan por la ausencia de rotación factorial. El AFC es, en realidad, un caso especial de un conjunto más general de métodos de análisis estadístico que se conocen como *análisis estructural de covarianza*. En esta sección se proporciona una introducción al AFC y se muestra cómo difiere del AFE.

Según diversos autores, en su forma actual, el AFC se atribuye a Joreskog y sus colegas (Joreskog, 1967, 1969, 1970; Joreskog y Goldberger, 1972), aunque Bock y Bargmann (1966) sugirieron algo similar en una fecha anterior. El procedimiento de Beck y Bargmann requiere que el investigador especifique todos los parámetros en la matriz de cargas factoriales, la matriz de correlaciones entre los factores y la matriz de varianza única. Con el empleo de tales matrices especificadas previamente, se crea una matriz de correlación estimada o de covarianza. Después esta matriz se compara con la correlación muestra o matriz de covarianza por medio de un estadístico de bondad de ajuste, tal como la prueba de Bartlett (véase Morrison, 1967). Dicho procedimiento por lo general resulta difícil de llevar a cabo, a menos que el investigador sepa qué valores iniciales de la matriz serían apropiados. Bernstein (1988) se refiere a lo anterior como la "solución forzada débil".

Joreskog (1969) reconoció que se podían estimar tan sólo algunos de los parámetros, pero no todos. El desarrollo de Joreskog permite al investigador especificar algunos de los parámetros y permite que los otros se estimen a partir de los datos. Estas modificaciones hacen del método de Joreskog un importante avance respecto de los procedimientos previos. Los cálculos para efectuar un análisis factorial confirmatorio se realizan por medio de un programa computacional. En la actualidad, los programas de elección son LISREL, EQS y AMOS, los cuales se utilizan para el análisis estructural de covarianza. Cada uno usa un algoritmo ligeramente distinto para realizar los cálculos. Con la aparición de cada nuevo programa, se facilita más su uso.

No obstante, como con el AFE, el uso apropiado de métodos de AFC como el LISREL, EQS y AMOS requiere que el investigador posea un conocimiento amplio del área que se estudia y de lo que representa una "buena" hipótesis respecto a la estructura factorial subyacente. No se trata de un método que pueda aplicarse exitosamente a un gran cuerpo de datos con muchas variables, donde el investigador no tiene idea alguna sobre cuál puede ser la estructura factorial subyacente.

Anteriormente en este capítulo se explicó la matriz de correlación. Dicha matriz o tabla contiene el coeficiente de correlación de cada variable con cada una de las otras variables. En notación de matriz lo anterior se escribe **R**. Para comprender el AFC, resulta necesario volver a examinar la ecuación fundamental del análisis factorial presentada anteriormente:

$$\mathbf{R} = \mathbf{P}\,\mathbf{\Phi}\,\mathbf{P}^{\mathrm{T}} + \mathbf{U}$$

La matriz o tabla **P** contiene las cargas factoriales. También se conoce como la matriz de patrón factorial. Si se vuelve a escribir esta matriz, de tal manera que sus renglones estén

intercambiados con sus columnas, dicha matriz transformada se denominaría la transposición de **P**, y se escribiría $\mathbf{P}^{\mathbf{T}}$. La matriz **Φ** indica qué tanto están correlacionados los datos entre sí. **U** representa la cantidad de singularidad dentro de cada variable.

La meta consiste en encontrar los valores de **P**, **Φ** y **U** que mejor reproduzcan la matriz **R** de correlación. La matriz de correlación reproducida se simboliza **R**′. Por ende, se tiene la ecuación de análisis factorial:

$$\mathbf{R}' = \mathbf{P}\,\Phi\,\mathbf{P}^{\mathbf{T}} + \mathbf{U}$$

Una solución aceptable para **P**, **Φ** y **U** sería aquella donde **R**′ y **R** difieran por una muy pequeña cantidad. En otras palabras, los valores encontrados dentro de estas matrices son tales que **R** y **R**′ tienen un buen ajuste. Uno de los índices más populares de bondad de ajuste es el índice de ajuste normado de Bentler y Bonnet (1980). Existe una cantidad de estadísticos de bondad de ajuste que se emplean regularmente (véase Comrey y Lee, 1992, capítulo 12).

Para desarrollar el modelo del AFC, el investigador debe elegir cuáles valores dentro de las matrices **P**, **Φ** y **U** se van a fijar y cuáles se van a estimar. Se pueden imponer restricciones sobre ciertos valores, tales como especificar un rango de valores dentro de los cuales deben de caer. Dichos valores también se denominan *parámetros*.

Los modelos de ecuación estructural para el análisis factorial confirmatorio, tal como se implementan por medio del LISREL y EQS, son nuevos procedimientos analíticos factoriales poderosos que con frecuencia representan el método de elección en una situación dada. Sin embargo, no son los únicos modelos que podrían emplearse, y en muchos casos se preferirían otros métodos.

Cualquiera que sea el método que los investigadores encuentren más apropiado para sus datos, resulta claro que el análisis factorial ha experimentado una revolución importante en los años recientes. Actualmente están disponibles nuevos métodos poderosos que expanden de manera formidable la capacidad de los trabajadores de investigación para examinar las implicaciones de sus datos. No obstante, debe enfatizarse que los métodos más nuevos deben considerarse como complementarios más que como sustitutos de los antiguos métodos del AFE, los cuales deben continuar siendo los procedimientos más efectivos para enfrentar situaciones de análisis con muchos datos. Por consiguiente, en el futuro previsible los estudiantes del análisis factorial necesitarán familiarizarse tanto con el método AFE como con el método AFC.

Ejemplo de investigación usando el análisis factorial confirmatorio

El capítulo de Keith (1997) brinda diversos ejemplos de investigación donde el AFC se aplica a problemas dentro de la psicología escolar. Este sobresaliente artículo es de fácil lectura, está bien escrito y es muy recomendable para quienes buscan buenos ejemplos donde se utilice el análisis factorial confirmatorio. Keith prueba la estructura de diversas

TABLA 34.7 *Estructura teórica del PIAK*

Inteligencia fluida	Inteligencia cristalizada	Recuerdo retardado
Aprendizaje de claves (rebus) visual-fonéticas	Definiciones	Evocación retrasada de claves (rebus)
Pasos lógicos	Comprensión auditiva	Evocación auditiva retardada
Códigos misteriosos	Significados dobles	
Memoria para diseño con cubos	Rostros famosos	

pruebas psicológicas para determinar si sus pretensiones son verdaderas, por medio del uso del análisis factorial confirmatorio. Él ofrece una explicación completa del modelo que va a probarse y de los estadísticos de bondad de ajuste proporcionados por el análisis factorial confirmatorio. En una demostración como ésta, Keith probó las pretensiones de la *prueba de inteligencia para adultos de Kaufman* (PIAK) (Kaufman Adult Intelligence Test), que consiste de 10 subescalas. Cuatro de ellas están diseñadas para medir la parte fluida de la inteligencia (*gf*); otras cuatro miden la parte cristalizada de la inteligencia (*gc*), y dos pruebas fueron diseñadas para medir la *evocación retardada*. Los componentes fluido y cristalizado de la inteligencia se mencionaron anteriormente cuando se explicó la teoría de Cattell sobre la inteligencia, la cual posteriormente fue modificada por John Horn (véase Cattell, 1987; Horn y Cattell, 1966) y ahora se llama teoría de la inteligencia de Horn-Cattell. Las partes cristalizada y fluida de la inteligencia constituyen sólo dos partes de la teoría completa. La evocación retardada mide la memoria de material aprendido en las primeras partes de la prueba, de quien responde la prueba. La tabla 34.7 presenta la estructura teórica o constructo del PIAK.

Con la ecuación fundamental del análisis factorial $\mathbf{R} = \mathbf{P}\,\Phi\,\mathbf{P}^T + \mathbf{U}$, los parámetros dentro de estas matrices pueden diseñarse de la siguiente manera:

$$\Phi = \begin{array}{c} \\ F1 \\ F2 \\ F3 \end{array} \begin{array}{c} \begin{array}{ccc} F1 & F2 & F3 \end{array} \\ \begin{bmatrix} 1.00 & & \\ * & 1.00 & \\ * & * & 1.00 \end{bmatrix} \end{array}$$

$$P = \begin{array}{c} \\ X_1 \\ X_2 \\ X_3 \\ X_4 \\ X_5 \\ X_6 \\ X_7 \\ X_8 \\ X_9 \\ X_{10} \end{array} \begin{array}{c} \begin{array}{ccc} F1 & F2 & F3 \end{array} \\ \begin{bmatrix} * & 0 & 0 \\ * & 0 & 0 \\ * & 0 & 0 \\ * & 0 & 0 \\ 0 & * & 0 \\ 0 & * & 0 \\ 0 & * & 0 \\ 0 & * & 0 \\ 0 & 0 & * \\ 0 & 0 & * \end{bmatrix} \end{array}$$

$$U = \begin{array}{c} \\ X_1 \\ X_2 \\ X_3 \\ X_4 \\ X_5 \\ X_6 \\ X_7 \\ X_8 \\ X_9 \\ X_{10} \end{array} \begin{array}{c} \begin{array}{cccccccccc} X_1 & X_2 & X_3 & X_4 & X_5 & X_6 & X_7 & X_8 & X_9 & X_{10} \end{array} \\ \begin{bmatrix} * & & & & & & & & & \\ 0 & * & & & & & & & & \\ 0 & 0 & * & & & & & & & \\ 0 & 0 & 0 & * & & & & & & \\ 0 & 0 & 0 & 0 & * & & & & & \\ 0 & 0 & 0 & 0 & 0 & * & & & & \\ 0 & 0 & 0 & 0 & 0 & 0 & * & & & \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & * & & \\ * & 0 & 0 & 0 & 0 & 0 & 0 & 0 & * & \\ 0 & 0 & 0 & 0 & 0 & * & 0 & 0 & 0 & * \end{bmatrix} \end{array}$$

Los asteriscos en cada matriz indican los parámetros que se van a estimar por medio de los datos. En la matriz Φ los asteriscos son las correlaciones entre los factores. En la matriz **P**, los asteriscos son las cargas factoriales; y en la matriz **U** los asteriscos representan las correlaciones entre las variables. En la matriz Φ los valores de la diagonal se fijan en 1.00. En la matriz **P** la solución deseada tendría una estructura más simple. Las variables sin marca están forzadas a tener valores de cero. En la matriz **U** existe una varianza única para cada una de las 10 variables. También tienen asteriscos debido a que serán estimadas por medio de los datos. En este modelo en particular, Keith afirma que el *aprendizaje de claves* (*rebus*) *visual-fonéticas* (variable 1) puede correlacionarse con la *evocación retardada de claves* (*rebus*) *visual-fonéticas* (variable 9) y que la *evocación auditiva retardada* (variable 10) estaría relacionada con la *comprensión auditiva* (variable 6). Tales correlaciones también se estiman a partir de los datos y, por lo tanto, dichos valores reciben asteriscos en la matriz **U**. Cuando se habla de los "datos", se refiere a las puntuaciones de los participantes en las 10 variables y a la matriz de intercorrelaciones para esas 10 variables.

Una vez establecido el modelo y los parámetros a estimar, es posible escribir los comandos de control adecuados para programas computacionales como el EQS, LISREL y AMOS. Los problemas de esta naturaleza son demasiado laboriosos para resolverse a mano.

Keith (1997) presenta el modelo y los estimados en forma de diagrama de ruta, y éstos como se vio anteriormente, son modelos conceptuales y visuales útiles para problemas en el análisis factorial confirmatorio y en el modelamiento de ecuación estructural.[1]

Keith encontró que los estadísticos de bondad de ajuste indican que el modelo PIAK se ajusta a los datos observados; además presenta seis de dichos estadísticos de bondad de ajuste y ofrece una explicación sobre ellos. La prueba chi cuadrada que se ha revisado en capítulos previos sirve como un estadístico de bondad de ajuste; sin embargo, se ve afectada por cambios en el tamaño de la muestra. Existen otros que son más adecuados.

El anterior representa sólo uno de dichos modelos demostrados por Keith. Su artículo continúa demostrando diversas variaciones distintas de modelos que llegan a comprobarse. Keith afirma que hay muchos más modelos y problemas que el AFC es capaz de realizar. De la misma manera en que Comrey y Lee (1992) demuestran cómo probar una estructura factorial hipotética en dos muestras separadas de manera simultánea, Keith indica cómo probar las similitudes de los factores a través de diferentes pruebas de inteligencia; por ejemplo, la *escala Wechsler de inteligencia para niños* (o WISC) y la *batería de Kaufman de evaluación para niños* (o K-ABC).

Análisis factorial e investigación científica

El análisis factorial tiene dos propósitos básicos: explorar áreas de variables para identificar los factores que presuntamente subyacen a las variables, y, como en todo trabajo científico, probar hipótesis sobre las relaciones entre variables. El primer propósito es muy reconocido y bastante bien aceptado. El segundo propósito no es tan reconocido ni tan aceptado.

Para conceptualizar el primer propósito —el exploratorio o reductivo— se debe tener en cuenta la validez de constructo y las definiciones constitutivas. El análisis factorial se concibe como una herramienta para la validez de constructo. Recuerde que en el capítulo 28 la validez se definió como la varianza del factor común. Puesto que la principal preocupación del análisis factorial es la varianza del factor común, por definición está firmemente relacionada con la teoría de medición. De hecho, tal relación fue expresada anteriormente

[1] El modelamiento de la ecuación estructural es un conjunto de términos alternativos para el análisis estructural de covarianza.

en la sección denominada "Un poco de teoría factorial", donde se anotaron las ecuaciones para aclarar la teoría analítica factorial. (Véase, en especial, la ecuación 34.6.)

Recuerde también que la validez de constructo busca el "significado" de un constructo a través de las relaciones entre el constructo y otros constructos. En capítulos iniciales de la presente obra, cuando se explicaron los tipos de definiciones, se aprendió que los constructos podían definirse de dos maneras: por medio de definiciones operacionales, y a través de definiciones constitutivas, las cuales son aquellas que definen constructos con otros constructos. En esencia, lo anterior es lo que hace el análisis factorial. Puede denominársele como un método de significado constitutivo, ya que permite al investigador estudiar los significados constitutivos de los constructos y, por consiguiente, su validez de constructo.

Las medidas de tres variables, por ejemplo, pueden tener algo en común. Este algo es, en sí mismo, una variable, presuntamente una entidad más básica que las variables utilizadas para aislarla e identificarla. A dicha nueva variable se le da un nombre; en otras palabras, se construye una entidad hipotética. Entonces, para indagar sobre la "realidad" de la variable es posible diseñar sistemáticamente una medida de ella y probar su "realidad" correlacionando los datos obtenidos con la medida con los datos de otras medidas teóricamente relacionadas con ella. El análisis factorial ayuda a verificar las expectativas teóricas.

Parte de los aspectos básicos de la vida de cualquier ciencia son sus constructos. Continúan usándose los constructos antiguos y constantemente se inventan nuevos. Note algunos de los constructos generales que son directamente pertinentes a la investigación educativa y del comportamiento: el aprovechamiento, la inteligencia, el aprendizaje, las aptitudes, las actitudes, la habilidad de solución de problemas, las necesidades, los intereses, la creatividad, el conformismo. Ahora considere algunas de las variables más específicas, importantes en la investigación del comportamiento: la ansiedad ante las pruebas, la habilidad verbal, el tradicionalismo, el pensamiento convergente, el razonamiento aritmético, la participación política y la clase social. Claramente una gran porción del esfuerzo de la investigación científica del comportamiento debe ser dedicado a lo que podría denominarse *investigación del constructo* o *validación del constructo*, y ello requiere del análisis factorial.

Cuando aquí se habla de relaciones, se refiere a las relaciones entre constructos: inteligencia y aprovechamiento, autoritarismo y etnocentrismo, reforzamiento y aprendizaje, clima organizacional y desempeño administrativo: todas las cuales son relaciones entre constructos demasiado abstractos o variables latentes. Dichos constructos, por lo común, deben definirse operacionalmente para poder estudiarse. Los factores son variables latentes, por supuesto, y el principal esfuerzo analítico factorial científico en el pasado se ha realizado para identificar factores y para utilizarlos ocasionalmente para medir variables de investigación. Raras ocasiones se han llevado a cabo intentos deliberados para evaluar los efectos de variables latentes sobre otras variables. No obstante, con los recientes avances y desarrollos en el pensamiento y metodología multivariados, resulta claro que ahora es posible evaluar la influencia de las variables latentes entre sí. Dicho desarrollo importante se analizará e ilustrará en el capítulo 35 sobre el análisis de estructuras de covarianza. Ahí se verá que los científicos llegan a obtener índices de las magnitudes y significancia estadística sobre los efectos de variables latentes en otras variables latentes. Si así ocurre, entonces, el análisis factorial se vuelve aún más importante en la identificación de variables o factores latentes, y el científico debe tener mucho cuidado con la interpretación de los datos, en los cuales se está evaluando la influencia de variables latentes.

Entonces, muchas áreas de investigación muy bien pueden ir precedidas por explotaciones analíticas factoriales de las variables del área. Lo anterior no significa que se junte una cantidad de pruebas y se aplique a cualquier muestra que parezca estar disponible. Las

investigaciones analíticas factoriales, tanto exploratorias como de comprobación de hipótesis, deben plantearse cuidadosamente. Es necesario controlar las variables que tengan alguna influencia: sexo, educación, clase social, inteligencia, etcétera. Las variables no se incluyen en un análisis factorial tan sólo por incluirlas. Deben tener un propósito legítimo. Si, por ejemplo, no es posible controlar la inteligencia por medio de la selección de la muestra, se incluye una medida de inteligencia (verbal quizás) en la batería de medidas. Al identificar la varianza de la inteligencia, en cierto sentido se controla la inteligencia. Se puede saber si las propias medidas están contaminadas por sesgos de respuesta, al incluir medidas del sesgo de respuesta, en el análisis factorial.

El segundo propósito principal del análisis factorial es la prueba de hipótesis. Ya se sugirió un aspecto de la comprobación de hipótesis: es factible incluir pruebas o medidas dentro de baterías analíticas factoriales de forma deliberada, para probar la identificación y naturaleza de los factores. El diseño de dichos estudios fue bien establecido por Comrey, Thurstone, Cattell, Guilford y otros. Primero, los factores son "descubiertos". Se infiere su naturaleza a partir de las pruebas que están cargadas en ellos. Dicha "naturaleza" se establece como una hipótesis. Se construyen y se aplican nuevas pruebas con nuevas muestras de sujetos. Los datos se analizan factorialmente. Si los factores surgen tal como *se predijo*, la hipótesis, hasta este punto, se confirma; parecería que los datos poseen "realidad". Pero ciertamente con ello no termina el asunto. Aún se deben probar, entre otras cosas, las relaciones de los factores con otros factores. Aún se deben ubicar los factores, como constructos, en una red nomológica de constructos.

Un uso menos reconocido del análisis factorial, descrito por Fruchter (1966), implica la comprobación de hipótesis experimentales. Por ejemplo, se hipotetiza que cierto método de enseñanza de lectura cambia los patrones de habilidad de los alumnos, de tal manera que la inteligencia verbal no es una influencia tan poderosa como lo es en otros métodos de enseñanza. Se puede planear un estudio experimental para probar esta hipótesis. Los efectos de los métodos de enseñanza se evalúan por medio de los análisis factoriales de un conjunto de pruebas aplicadas antes y después del uso de los diferentes métodos. Woodrow (1938) comprobó una hipótesis similar cuando aplicó un conjunto de pruebas antes y después de la práctica en siete pruebas: sumar, restar, anagramas, etcétera. Encontró que los patrones de carga factorial *sí* cambiaron después de la práctica.

Al considerar el valor científico del análisis factorial debe prevenirse al lector de no atribuir "realidad" y singularidad a los factores. El riesgo de materialización es muy grande. Resulta sencillo nombrar un factor y después creer que existe una realidad detrás del nombre. Sin embargo, el hecho de asignarle un nombre a un factor no le confiere realidad. Los nombres de los factores son meros intentos de comprender la esencia de los factores. Siempre son tentativos, sujetos a confirmación o desconfirmación posterior. Asimismo, muchas cosas pueden producir factores. Cualquier asunto que introduzca correlación entre variables "crea" un factor. Las diferencias en sexo, educación, antecedentes sociales y culturales y en inteligencia, quizá provoquen la aparición de factores. Los factores también difieren —por lo menos hasta cierto punto— en muestras diferentes. Conjuntos de respuestas o formas de pruebas pueden hacer surgir factores. Aun con estas precauciones, debe señalarse que los factores sí surgen repetidamente en diferentes pruebas, diferentes muestras y diferentes condiciones. Cuando así sucede, se tiene bastante certeza de que existe una variable subyacente que se está midiendo exitosamente.

Como se mencionó al inicio del capítulo, existen críticas serias al análisis factorial. Las principales críticas válidas se centran alrededor de la indeterminación de cuántos factores se deben extraer de una matriz de correlación, y del problema de cómo rotar los factores. Otra dificultad que molesta tanto a los críticos como a los adeptos es lo que puede llamarse el "problema de la comunalidad", o qué cantidades poner dentro de la diagonal de la

matriz R antes de factorizar. En un capítulo introductorio estos problemas no pueden analizarse en detalle. Se refiere al lector a la explicación de Cattell (1978), Comrey y Lee (1992), Cureton y D'Agostino (1983), Gorsuch (1983), Guilford (1954), Harman (1976) y Thurstone (1947). Una crítica de otro tipo parece inquietar a los educadores y sociólogos, así como a algunos psicólogos. Ésta adquiere dos o tres formas que parecen reducirse en desconfianza, algunas veces profunda, combinada con antipatía hacia el método, a causa de su complejidad y, de manera extraña, a su objetividad.

El argumento expresa algo como esto. El análisis factorial junta demasiadas pruebas dentro de una máquina estadística y arroja factores que tienen poco significado psicológico o sociológico. Los factores son meros artefactos del método. Son promedios que no corresponden a realidad psicológica alguna, especialmente a la realidad psicológica del individuo, que no sea otra que la que está en la mente del analista factorial. Además, no se puede obtener más del análisis factorial de lo que se haya puesto dentro de él.

El argumento se vuelve básicamente irrelevante. Decir que los factores no tienen significado psicológico y que son promedios es tanto verdadero como falso. Si los argumentos fueran válidos, ningún constructo científico tendría significado alguno. Todos son, en cierto sentido, promedios. Todos son inventos del científico. Se trata sencillamente del terreno de la ciencia. El criterio básico de la "realidad" de cualquier constructo, de cualquier factor, es su "realidad" empírica, científica. Si, después de descubrir un factor, se predicen exitosamente relaciones a partir de presuposiciones teóricas e hipótesis, entonces el factor posee "realidad". No existe mayor realidad en un factor que ésta, de la misma forma que no existe mayor realidad en un átomo que sus manifestaciones empíricas.

El argumento que sostiene que sólo se obtiene lo que se pone en un análisis factorial no tiene significado alguno y también es irrelevante. Ningún investigador competente del análisis factorial afirmaría algo más que esto. Pero esto no quiere decir que nada se descubra en el análisis factorial. Todo lo contrario. La respuesta es, por supuesto, que no se obtiene nada más del análisis factorial que lo que se pone en él, pero que no se conoce *todo* lo que se pone en él. Tampoco se conoce cuáles pruebas o medidas comparten varianza del factor común; tampoco se conocen las relaciones entre los factores. Únicamente el estudio y el análisis llegan a indicar estas cosas. Se podría desarrollar una escala de actitud que se piense mide una sola actitud. Naturalmente, un análisis factorial de los reactivos de actitud no genera factores que no estén en los reactivos. Sin embargo, puede mostrar, por ejemplo, que existen dos o tres fuentes de varianza común en una escala que se creía unidimensional. De manera similar, una escala que se creía que estaba midiendo *autoritarismo* puede mostrar, por medio del análisis factorial, que mide *inteligencia*, *dogmatismo* y otras variables.

Si se examina la evidencia empírica más que la opinión, se debe concluir que el análisis factorial constituye una de las herramientas más poderosas diseñadas hasta ahora para el estudio de áreas complejas de interés científico del comportamiento. De hecho, el análisis factorial es uno de los inventos creativos del siglo XX, así como lo son las pruebas de inteligencia, el condicionamiento, la teoría del reforzamiento, la definición operacional, el concepto de aleatoriedad, la teoría de medición, el diseño de investigación, el análisis multivariado, la computadora y las teorías del aprendizaje, la personalidad, del desarrollo, de organizaciones y de la sociedad.

Es adecuado que el capítulo se concluya con algunas palabras de un gran científico, maestro y analista factorial de la psicología, Louis Leon Thurstone (1959, p. 8):

> Como científicos, tenemos fe en que las habilidades y personalidades de la gente no sean tan complejas como la enumeración total de los atributos que pueden listarse. Creemos que estas características se componen de un menor número de factores o elementos pri-

marios que se combinan de varias maneras para formar una larga lista de características. Es nuestra ambición encontrar algunas de dichas habilidades y características elementales.

Todo trabajo científico tiene algo en común: que se intenta comprender la naturaleza de la forma más parsimoniosa. Una explicación de un conjunto de fenómenos o un conjunto de observaciones experimentales logra aceptación sólo en tanto ofrezca control intelectual o comprensión de una variedad relativamente amplia de fenómenos, en términos de un número limitado de conceptos. El principio de la parsimonia es intuitivo para cualquiera que tenga aun la más leve aptitud para la ciencia. La motivación fundamental de la ciencia es el anhelo de la comprensión más simple posible de la naturaleza, y encuentra satisfacción en el descubrimiento de las uniformidades simplificadoras llamadas leyes científicas.

RESUMEN DE CAPÍTULO

1. El análisis factorial examina un conjunto de variables y determina cuáles van juntas. Las variables que se agrupan se denominan *un factor*.
2. Un factor es un constructo, una entidad hipotética o una variable latente que fundamenta las mediciones de cualquier tipo.
3. Charles Spearman desarrolló el análisis factorial, pero fue Louis Thurstone quien lo expandió y mejoró. Thurstone se considera el "padre del análisis factorial moderno".
4. Algunas de las contribuciones de Thurstone incluyen el método centroide de extracción, la rotación factorial y la estructura simple. Tanto la rotación como la estructura simple se emplean para volver más interpretables los factores.
5. La ecuación fundamental del análisis factorial es $\mathbf{R} = \mathbf{P}\,\Phi\,\mathbf{P}^{T} + \mathbf{U}$.
6. La ecuación fundamental muestra cómo la matriz de correlación **R** se parte en una matriz **P** de carga factorial, correlación entre factores, Φ, y singularidad, **U**. **R** son los datos observados. Todas las demás se estiman a partir de los datos.
7. Existe una variedad de métodos de extracción factorial diferentes. El más popular ha sido el método de factores principales.
8. El método de factores principales requiere que el investigador aporte estimados de las comunalidades y que establezca el número de factores a extraer.
9. Los estimados de las comunalidades y el número de factores han sido problemas difíciles de resolver en el análisis factorial. No existe un conjunto de reglas claras para cada uno. Sin embargo, Comrey ha desarrollado un método que no utiliza estimados de las comunalidades.
10. La comunalidad se refiere a la proporción de la varianza total que es varianza del factor común. Una meta del análisis factorial consiste en encontrar los componentes de varianza de la varianza de factor común total.
11. Las puntuaciones factoriales implican la combinación de los valores de aquellas variables que definan al factor. Es una puntuación transformada nueva. Si una batería de 10 pruebas produce tres factores, entonces cada persona que responde las pruebas tendría puntuaciones de tres factores.
12. Actualmente el análisis factorial sigue dos metodologías. El método tradicional ahora se llama *análisis factorial exploratorio* o AFE, y el método más nuevo se denomina *análisis factorial confirmatorio* o AFC.
13. El análisis factorial exploratorio por lo común se utiliza para comprender o descubrir cuáles factores subyacen a los datos. Algunos investigadores que son usuarios experimentados de este método saben cómo probar hipótesis sobre los factores.

14. El análisis factorial confirmatorio sirve para probar hipótesis acerca de la estructura factorial. En el AFC se desarrolla un modelo, basado en la teoría o en hallazgos previos, y luego se prueba contra los datos empíricos.
15. El análisis factorial confirmatorio es sólo un caso especial de un grupo de análisis llamados *análisis estructural de covarianza*.

SUGERENCIAS DE ESTUDIO

1. El estudiante más avanzado encontrará valiosa la siguiente selección de artículos:

Comrey, A. L. (1978). Common methodological problems in factor analitic studies. *Journal of Consulting and Clinical Psychology, 46*, 648-659. [Una revisión no matemática de los problemas encontrados en estudios analíticos factoriales.]

Comrey, A. L. (1985). A method for removing outliers to improve factor analytic results. *Multivariate Behavioral Research, 20*, 273-281. [Muestra cómo detectar y eliminar valores extremos que ejercen un efecto negativo sobre la solución de un análisis factorial.]

Comrey, A. L. y Montag, I. (1982). Comparison of factor analytic results with two-choice and seven-choice personality item formats. *Applied Psychological Measurement, 6*, 285-289. [Comparación de dos resultados de un análisis factorial de las *escalas de personalidad de Comrey*, donde uno de los análisis se realizó con un formato de respuesta de dos opciones, y otro con un formato de siete opciones. Los resultados indican la superioridad del formato de siete opciones sobre el de dos opciones, para los inventarios de personalidad.]

Dunlap, W. P. y Cornwell, J. M. (1994). Factor analysis of ipsative measures. *Multivariate Behavioral Research, 29*, 115-126. [Explora el análisis factorial con medidas ipsativas. Los autores muestran de forma analítica los problemas fundamentales que las medidas ipsativas imponen al análisis factorial. Tales investigadores recomiendan que el análisis factorial no debe realizarse con datos que sean reconocidos como ipsativos.]

Fleming, J. S. (1981). The use and misuse of factor scores in multiple regression analysis. *Educational and Psychological Measurement, 41*, 1017-1025. [Explora cuándo y dónde puede usarse el análisis factorial junto con la regresión múltiple con propósitos predictivos.]

Lee, H. B. y Comrey, A. L. (1979). Distortions in a commonly used factor analytic procedure. *Multivariate Behavioral Research, 14*, 301-321. [Un estudio que compara el método más popular de extracción y rotación factorial con otros métodos. Muestra cuán distorsionadas se pueden volver algunas soluciones al utilizar los métodos más populares.]

Montanelli, R. y Humphreys, L. (1976). Latent roots of random data correlation matrices with squared multiple correlations on the diagonal: A Monte Carlo study. *Psychometrika, 41*, 341-348. [Excelente método de correlación y regresión aleatorias respecto al problema del número de factores.]

Overall, J. (1965). Note on the scientific status of factors. *Psychological Bulletin, 61*, 270-276. [Un análisis excelente e incluso brillante, sobre nociones básicas del análisis factorial.]

Peterson, D. (1965). Scope and generality of verbally defined personality factors. *Psychological Review, 72*, 48-59. [Muy convincente sobre el problema del número de factores.]

2. Como siempre, no existe un sustituto para el estudio de los usos de los métodos en la investigación real. Por lo tanto, el lector debe consultar dos o tres buenos estudios sobre el análisis factorial. Selecciónelos de los citados en el capítulo o de los siguientes:

   Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies.* Nueva York: Cambridge University Press. [Un análisis y reanálisis de estudios de inteligencia que utilizaron análisis factorial. Ofrece una muy buena historia de psicología de las diferencias individuales.]

   Daniel, L. G. y Siders, J. A. (1994). Validation of teacher assessment instruments: A confirmatory factor analytic approach. *Journal of Personnel Evaluation in Education, 8,* 29-40. [Examen de la validación de constructo del Mississippi Teacher Assessment Instrument utilizado para la certificación de nuevos maestros. Un análisis factorial exploratorio encontró cuatro factores; sin embargo, análisis factoriales confirmatorios no lograron generar un modelo estructural aceptable.]

   Fleming, J. S. y Whalen, D. J. (1990). The personal and academic self-concept inventory: Factor structure and gender differences in high school and college samples. *Educational and Psychological Measurement, 50,* 957-967. [Análisis factorial confirmatorio aplicado a diversos modelos estructurales competentes, del Personal and Academic Self-Concept Inventory, una extensión de las Self-Rating Scales.]

   Isaacson, R. L. McKeachie, W. L., Milholland, J. E. y Lin, Y. G. (1964). Dimensions of student evaluations of teaching. *Journal of Educational Psychology, 55,* 344-351. [Un estudio competente de los factores que subyacen a las evaluaciones de los estudiantes respecto a los instructores. El primer factor es importante.]

   Mitrushina, M. y Satz, P. (1991). Changes in cognitive functioning associated with normal aging. *Archives of Clinical Neuropsychology, 6,* 49-60. [Uso del análisis factorial en pruebas de memoria y psicomotrices para encontrar factores de funcionamiento cognitivo en participantes ancianos. Se calcularon las puntuaciones factoriales y se utilizaron en un análisis de varianza.]

   Thurstone, L. L. (1944). *A factorial study of perception. Psychometric Monographs,* núm. 4. Chicago: University of Chicago Press. [Otro estudio pionero y clásico de Thurstone.]

3. A continuación se presenta una pequeña matriz de correlación ficticia, con los nombres de las pruebas.

		1	2	3	4	5	6
1.	Vocabulario	.70	.22	.20	.15	.25	
2.	Analogías	.70		.15	.26	.12	.30
3.	Suma	.22	.15		.81	.21	.10
4.	Multiplicación	.20	.26	.81		.31	.29
5.	Recuerdo de nombres propios	.15	.12	.21	.31		.72
6.	Reconocimiento de figuras	.25	.30	.10	.29	.72	

   a) Realice un análisis factorial a simple vista. Es decir, por medio de la inspección de la matriz determine cuántos factores probablemente hay y qué pruebas están en qué factores.

b) Asigne un nombre a los factores. ¿Qué tan seguro está de sus nombres? ¿Qué puede hacer usted para estar más seguro de sus conclusiones?

4. Algunos excelentes libros sobre el análisis factorial que uno desearía leer para obtener información son los siguientes:

Cattell, R. B. (1978). *The scientific use of factor analysis in the behavioral and life sciences.* Nueva York: Plenum. [Se trata de un libro sobresaliente que cubre todo lo que ha sucedido con el análisis factorial desde la publicación del libro de Cattell en 1952. Se compone de dos partes e introduce conceptos matemáticos de manera gradual.]

Comrey, A. L. y Lee, H. B. (1992). *A first course in factor analysis* (2a. ed.). Hillsdale, Nueva Jersey: Lawrence Erlbaum. [En lugar de introducir al estudiante a las matemáticas de las matrices en un capítulo, el libro presenta de forma gradual las matrices y su uso en el análisis factorial. Los temas en este libro suplementan los temas cubiertos en otros libros sobre el análisis factorial. El capítulo 11 resulta especialmente valioso, pues muestra al lector cómo se desarrollaron las *Escalas de Personalidad de Comrey* con el uso del análisis factorial. Se exponen métodos valiosos que normalmente no se encuentran en otros libros de texto.]

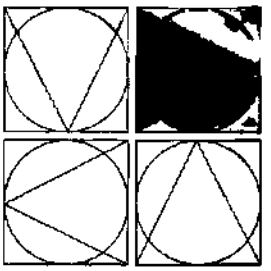
Cureton, E. E. y D'Agostino, R. B. (1983). *Factor Analysis: An applied approach.* Hillsdale, Nueva Jersey: Lawrence Erlbaum. [Un libro que presenta modelos y teorías del análisis factorial exploratorio, sin el empleo de matemáticas avanzadas como el cálculo. Contiene un buen capítulo sobre álgebra matricial y cubre adecuadamente el uso de métodos del principio del eje para el análisis factorial común. También proporciona algunas comparaciones entre diferentes métodos de extracción y rotación factoriales.]

Gorsuch, R. (1983). *Factor analysis* (2a. ed.). Hillsdale, Nueva Jersey: Lawrence Erlbaum. [Se trata de un libro académico, informativo y con autoridad. Una de sus grandes virtudes es que explora de forma profunda los problemas más difíciles y complicados del análisis factorial. Gorsuch no sólo explica las ideas técnicas, sino que también cita contribuciones teóricas e investigaciones empíricas de los problemas. *Sumamente recomendable* como trabajo de referencia para los investigadores del comportamiento.]

Mulaik, S. A. (1972). *The foundations of factor analysis.* Nueva York: McGraw-Hill. [Un tratamiento matemáticamente sofisticado del análisis factorial. Definitivamente se recomienda para quienes poseen un entrenamiento matemático extenso, como del cálculo multivariado. Este libro actualmente ya no se imprime pero puede estar disponible en algunas bibliotecas.]

Rummel, R. J. (1970). *Applied factor analysis.* Evanston, Illinois: Northwestern University Press. [Un libro profundo sobre el análisis factorial exploratorio con un énfasis en ciencias políticas. Requiere de conocimientos sobre matemáticas. Contiene buenos capítulos sobre álgebra matricial y su uso en la explicación del análisis factorial. Se proporciona una buena explicación sobre los diversos modelos factoriales. Sin embargo, no explica de forma extensa los asuntos implicados en el uso e interpretación de las soluciones analíticas factoriales.]

Thurstone, L. L. (1947). *Multiple factor analysis.* Chicago: University of Chicago Press. [Se trata de un trabajo clásico del creador del análisis factorial moderno. Aunque se escribió hace más de 50 años, el material que presenta aún es relevante.]



CAPÍTULO 35

ANÁLISIS ESTRUCTURAL DE COVARIANZA

En esta larga y complicada disertación sobre los fundamentos de la investigación del comportamiento, con frecuencia se ha hablado sobre la importancia de la teoría y de su comprobación. En ciertos momentos se ha enfatizado el propósito de la investigación científica de formular explicaciones de los fenómenos naturales y de someter las implicaciones de las explicaciones a una prueba empírica. En este capítulo se estudiará y se tratará de comprender un sistema analítico altamente desarrollado y conceptualmente sofisticado para modelar y probar teorías científicas del comportamiento: el *análisis estructural de covarianza*. El análisis estructural de covarianza también tiene otro nombre, algunas veces se denomina *modelamiento de ecuaciones estructurales* (MEE). Para entender esta metodología, se consideran de forma importante algunos sistemas matemático-estadísticos y programas computacionales. Existen al menos tres que son reconocidos, y que actualmente están en uso. A la mitad de los años setenta, el LISREL (*Linear Structural Relations*) (*relaciones estructurales lineales*) fue creado y desarrollado por Joreskog y sus colaboradores (Joreskog y Sorbom, 1993) para establecer y analizar estructuras de covarianza. Las primeras versiones de este programa computacional requerían del establecimiento de planteamientos difíciles. Sin embargo, las generaciones posteriores se hicieron mucho más fáciles. Hasta hace poco formaba parte del paquete estadístico SPSS y aún continúa siendo el método preferido para muchos diseñadores. A finales de los

setenta y principios de los ochenta, el programa computacional llamado EQS fue desarrollado por Bentler (1986). Los investigadores interesados en el análisis estructural de covarianza encontraron que el programa de Bentler resultaba más fácil de utilizar. Los planteamientos del programa y los símbolos del modelo eran más fáciles de comprender que los del LISREL. Sin embargo, el LISREL ha incrementado en mucho su número de usuarios con la aparición de la versión 8. Aunque ya es un poco antiguo, el trabajo de Brown (1986) comparó el LISREL y el EQS en términos de la estimación de parámetros para el análisis factorial confirmatorio. Aquí se utilizará el EQS debido a que muchos lo encuentran más fácil de entender, pues utiliza denominaciones estándar; mientras que el modelado LISREL emplea bastantes letras griegas. Sin embargo, una vez que el investigador está familiarizado con la estructura de covarianza o con el modelamiento de ecuación estructural, las diferencias se vuelven menos importantes.

Los investigadores tienen ahora un tercer sistema y programa computacional al cual enfrentarse: es el llamado análisis de estructuras momentáneas (Analysis of Moment Structures, AMOS), que fue publicado por SmallWaters Corporation (Arbuckle, 1995). Ellos tienen una versión de demostración de su programa en su sitio de Internet. La versión más reciente permite que el usuario especifique, vea y modifique el modelo estructural *gráficamente*, por medio del uso de herramientas de dibujo. Cada uno de dichos programas y modelos ha logrado, de manera consistente, que los investigadores trabajen con mayor facilidad el uso del modelamiento de la ecuación estructural o análisis estructural de covarianza.

Por desgracia, aun con programas computacionales mejorados y diversos manuales y libros sobre el tema (véase Schumacker y Lomax, 1996), no es fácil aprender el análisis estructural de covarianza o el modelamiento de ecuación estructural para quienes carecen de conocimientos matemáticos. Debe confesarse que la dificultad radica en explicar el sistema en lenguaje comprensible a aquellos que no saben leer las matemáticas y, al mismo tiempo, cumplir con los propósitos y objetivos de este libro. Por lo tanto, la exposición se limita a presentar y explicar el mero esqueleto matemático del sistema y a explicar cómo y por qué se utiliza. Por fortuna, el tema va íntimamente relacionado con las exposiciones del análisis de regresión múltiple y del análisis factorial de los capítulos 32, 33 y 34.

Estructuras de covarianza, variables latentes y comprobación de la teoría

El análisis estructural de covarianza puede considerarse como una combinación del análisis factorial y del análisis de regresión múltiple. De hecho, Lee y Jennrich (1984) han mostrado cómo emplear el análisis de regresión no lineal para analizar datos de estructuras de covarianza. Su ventaja más importante consiste en que se pueden evaluar los efectos de las variables latentes entre sí y sobre otras variables observadas. Recuerde que una *variable latente* es un constructo o "entidad" hipotética: inteligencia, destreza verbal, habilidad espacial, prejuicio, ansiedad, aprovechamiento. Las variables latentes son, por supuesto, variables no observadas, cuya "realidad" se asume o infiere a partir de variables o indicadores observados. Los factores son variables latentes, constructos que inventamos para explicar el comportamiento observado.

El análisis factorial confirmatorio se presentó en el capítulo 34, y éste constituye una forma de estructura de covarianza. El lector podrá recordar que un diagrama de ruta resultaba muy útil para conceptualizar la apariencia del modelo. En el análisis estructural de covarianza se utilizarán mucho los modelos de ruta. Una vez que el modelo de ruta se

construye de forma correcta, entonces puede usarse el EQS, LISREL o AMOS. A lo largo de este capítulo se utilizarán los modelos de ruta para describir el modelo de estructuras de covarianza.

Existen diez puntos clave que el diseñador del modelo debe considerar al dibujar el diagrama de ruta que se va a analizar con el uso del EQS. Si se siguen estos puntos, entonces el diagrama del análisis de ruta se ajustará a los planteamientos del programa EQS. Se listarán dichos puntos y después se explicará cada uno. Los 10 puntos son relevantes tanto para la estructura de covarianza más simple como para la más compleja.

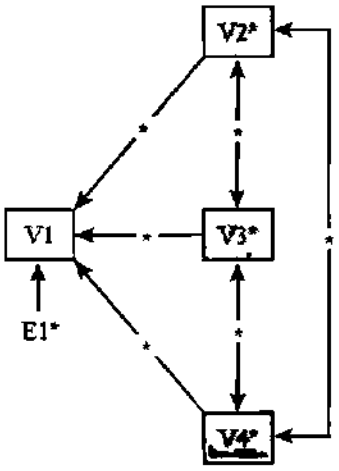
1. Existe una flecha unidireccional a partir de cada variable independiente, señalando hacia la variable dependiente.
2. Cada variable que tiene una flecha unidireccional apuntando hacia sí misma genera una ecuación de regresión lineal en el modelo de covarianza o de ecuación estructural.
3. Existe un asterisco (*) insertado en cada flecha, de las variables independientes a la variable dependiente, lo cual indica que existen parámetros libres a ser estimados para estas rutas.
4. El asterisco identifica un parámetro libre en el modelo.
5. Todas las covarianzas (correlaciones) entre las variables independientes también representan parámetros libres en el modelo. Los parámetros libres de la covarianza se indican por medio de flechas bidireccionales, con un asterisco en medio.
6. Las varianzas de las variables independientes medidas también son parámetros libres. Dichas variables se encuentran subrayadas en sus cajas, con un asterisco junto a su símbolo.
7. Todas las variables independientes poseen varianzas que funcionan como parámetros en el modelo.
8. Las variables dependientes no poseen varianzas que funcionen como parámetros en el modelo.
9. Todas las variables independientes latentes (sin medir) en el modelo deben tener su escala fija en una de dos maneras:
   a) Al establecer el coeficiente de regresión en un valor fijo. Por lo general, se establece en 1.0.
   b) Al fijar su varianza en algún valor conocido, generalmente 1.0.
10. En la mayoría de los casos, los valores E (error de medición) tienen sus coeficientes de regresión fijos en 1.0, y por ello no aparece ningún asterisco en la flecha que apunta hacia la variable dependiente.

En el EQS, el modelamiento requiere de variables independientes y dependientes. Cada una de ellas, o ambas, pueden ser medidas o latentes. En ocasiones, las variables latentes también se denominan *variables sin medir*. En el caso de las variables medidas, se utiliza el símbolo "E" para representar su error de medición. El símbolo "D" se utiliza para representar el error de medición de las variables dependientes latentes. Dependiendo de la numeración de las variables dependientes, aparece un número unido a la "E" o a la "D".

La estructura de covarianza más simple es el análisis de regresión. Si se establece que x_1 sea la variable dependiente, y que x_2, x_3 y x_4 sean las variables independientes, se escribe la ecuación del modelo de la siguiente manera:

$$x_1 = B_2x_2 + B_3x_3 + B_4x_4 + e \qquad (35.1)$$

FIGURA 35.1



Las puntuaciones en esta ecuación aparecen en forma de puntuación de desviación. Lo anterior vuelve innecesario el término de la intersección. Los valores de *B* son los pesos estandarizados de la regresión. El modelo de regresión de ecuación única mostrado en la ecuación 35.1 puede representarse con el diagrama de ruta de la figura 35.1.

Las variables de datos medidos están representadas en cajas o rectángulos e incluyen números de identificación como V1, V2, V3, V4 o más, según se requiera. Es decir, se utilizan cuadrados o rectángulos para encerrar una variable observada, y no latente. En el caso del análisis de regresión lineal, todas las variables se consideran observadas o medidas. Tal como se mencionó en el capítulo anterior, en la exposición sobre el análisis factorial confirmatorio, se utilizan círculos o elipses para encerrar variables latentes o sin medir. En el ejemplo de regresión existen cuatro variables en la ecuación: x_1, x_2, x_3, x_4. Por lo tanto, el EQS requiere que se nombren V1, V2, V3 y V4. Hay una flecha unidireccional a partir de cada una de las variables independientes V2, V3 y V4, que señalan hacia la variable V1. A partir del punto dos expresado antes, cada variable que tiene una flecha unidireccional apuntando hacia sí misma, genera una ecuación de regresión lineal en el modelo. Aquí sólo hay una variable de este tipo, V1 y, por ende, sólo una ecuación.

Existe un asterisco o estrella (*) incluida en cada una de las flechas que va de V2, V3 y V4 a V1, lo que indica que existen parámetros libres a estimarse en conexión con estas rutas, uno para cada estrella. También hay una flecha unidireccional apuntando desde E1 hacia V1, que indica que la ecuación de regresión contiene una variable de error, E1, que también es una variable independiente en el modelo. El asterisco (*) indica al programa que el valor precedente es un estimado y no un valor fijo. El asterisco también identifica un parámetro libre en el modelo. El conteo de los asteriscos permite al investigador determinar el número total de parámetros libres que se están estimando para el modelo. Se supone que E1 no está correlacionado con V2, V3 y V4; por lo tanto, no aparecen flechas bidireccionales entre estas variables, en el diagrama de ruta. Observe también que el coeficiente para E1 se estableció en 1.0. Se colocó el 1.0 para ser consistentes con el punto clave # 10. Por lo común, el 1.0 está implícito (no escrito).

Otros parámetros libres en el modelo incluyen todas las covarianzas entre las variables independientes medidas, V2, V3 y V4. Las flechas bidireccionales con un asterisco en medio indican parámetros de covarianza libres. Ello añade tres parámetros libres más.

Los parámetros libres adicionales incluyen las varianzas de las variables independientes medidas: V2, V3 y V4. El hecho de que tales varianzas sean parámetros libres se indica en el modelo con el sombreado de V2, V3 y V4 en las cajas, y por la aparición de asteriscos (*) junto con sus símbolos. Resulta sencillo perder de vista el hecho de que las varianzas de estas variables independientes son parámetros en el modelo. El sombreado facilita recordar que el esquema del programa EQS debe contener ya sea estimados o valores fijos para estos parámetros de varianza, lo cual añade tres parámetros libres más (indicados por los asteriscos).

Se debe estimar una varianza más en el modelo debido a que la regla general es que todas las variables independientes tienen varianzas que funcionan como parámetros en el modelo, y que incluyen todas las variables de *error*, así como las variables independientes *medidas*. Sin embargo, las variables dependientes no tienen varianzas que funcionen como parámetros en el modelo. Se colocó un asterisco (*) junto a E1 en el diagrama de ruta de la figura 35.1 para mostrar que su varianza es un parámetro libre en el modelo. Ahora esto da un total de diez asteriscos en el diagrama de ruta, lo cual indica que hay 10 parámetros libres por estimar.

Existe un parámetro adicional en el modelo que está fijo en 1.0 —el coeficiente de regresión para E1—. Redundando con el punto clave # 9, todas las variables independientes sin medir en el modelo deben fijar su escala en una de dos maneras: *a*) al establecer un coeficiente de regresión en un valor fijo, por lo general 1.0, que aquí se hizo para E1; o *b*) al fijar su varianza en un valor conocido, por lo general de 1.0. En la mayoría de los casos, los valores de E tienen sus coeficientes de regresión fijos en 1.0 y, como consecuencia, en el diagrama de ruta no aparece ningún asterisco (*) en la flecha que apunta hacia la variable dependiente, con la cual la variable de error esté asociada.

No es factible fijar tanto el peso de regresión *como* la varianza para una variable E, a causa de que el producto de ambos números debe estar libre para acomodar la cantidad de error, para predecir la variable dependiente de forma correcta. Por lo tanto, se fija uno u otro, pero no ambos.

Existen 10 parámetros libres a estimar en el modelo de regresión de la figura 35.1, a partir de un total de 10 puntos de datos. Los puntos de datos consisten en las varianzas y las covarianzas de las variables medidas, V1, V2, V3 y V4, o $(n(n + 1))/2$, donde n es el número de variables. Es decir, existen seis covarianzas y cuatro varianzas. El número de grados de libertad para la estimación del modelo está dado por el número de puntos de datos menos el número de parámetros libres en el sistema: en este caso, es 10 – 10 o cero.

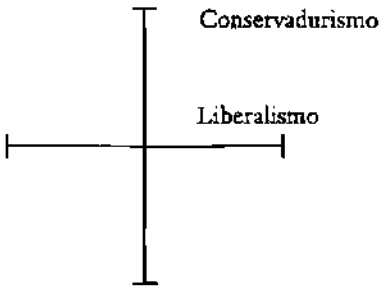
Cuando no existen grados de libertad, se dice que el modelo está "saturado"; es decir, es posible obtener valores para los parámetros libres, que reproducirán los datos de entrada de manera exacta. Por consiguiente, no existe duda sobre si el modelo se ajusta a los datos, y no se requiere de una prueba chi cuadrada ni de otra prueba estadística para saber qué tan bueno es el ajuste, pues el ajuste es perfecto. Por tal razón, los modelos de regresión no se consideran de mucho interés para el análisis estructural de covarianza. En general, quienes deciden utilizar el análisis estructural de covarianza buscan desarrollar un modelo que tenga considerablemente más puntos de datos que parámetros libres a estimar. En tales casos, se vuelve un reto encontrar un modelo no saturado y un conjunto de parámetros que reproduzcan los datos razonablemente bien, es decir, que den un buen ajuste. Sólo este tipo de modelo (uno con el número de grados de libertad mayor que cero) será capaz de ofrecer cualquier información científica con significancia teórica.

Si todo lo antes expuesto continúa pareciendo demasiado abstracto, ahora se examinará un ejemplo de una investigación real. Dicha investigación fue realizada por Kerlinger (1972). Posee las virtudes de la familiaridad y de la simpleza relativa.

Comprobación de hipótesis factoriales alternativas: dualidad contra bipolaridad de las actitudes sociales

Recuerde que existen dos perspectivas generales de las actitudes sociales que por lo general se asocian con el liberalismo y el conservadurismo. Una perspectiva —la más comúnmente adoptada por los científicos y la gente común— señala que los aspectos y la gente liberales y conservadores se oponen entre sí: el conservador está en contra de lo que el liberal apoya, y a la inversa. Esto se expresó antes como una *teoría bipolar*. Implica una dimensión de las actitudes, con los aspectos y la gente liberal en un extremo, y los aspectos y la gente conservadora en el otro.

La teoría, hipótesis o concepción contrastante sobre las actitudes sociales indica, en efecto, que los aspectos e ideas liberales son, en general, diferentes y virtualmente independientes de los aspectos e ideas conservadores. El liberalismo y el conservadurismo, para utilizar los nombres abstractos de las variables latentes, no necesariamente son opuestos entre sí: son dos ideologías separadas e independientes, o son conjuntos de creencias relacionadas que llegan a expresarse como dimensiones ortogonales:



Tal concepción de la estructura de las actitudes sociales es *dualista*.

Las dos "teorías" contrastantes de la estructura de las actitudes sociales pueden expresarse por medio de las dos matrices factoriales *A* y *B*, que se presentan en la tabla 35.1, las cuales pueden denominarse matrices "objetivo", debido a que se establecen para expresar análisis de contraste. Suponga que se han administrado seis escalas de actitudes sociales a una muestra grande heterogénea de individuos. Las escalas 1, 2 y 3 son escalas conservadoras, y las escalas 4, 5 y 6 son escalas liberales. Considere que las respuestas de la muestra a las seis escalas fueron correlacionadas y analizadas factorialmente. Los resultados del análisis factorial de lo que implican las teorías de dualidad y bipolaridad se muestran en la tabla. Los signos + indican cargas sustanciales y positivas, los signos – indican cargas sus-

TABLA 35.1 *Estructuras analíticas factoriales implicadas en la hipótesis dualista (A) y la hipótesis bipolar (B)**

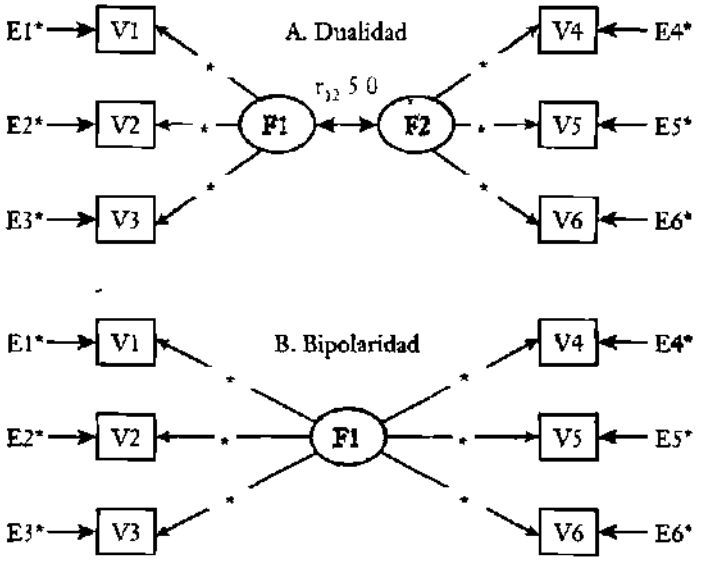
(A) Dualista Escalas	I	II	Tipo	(B) Escalas	Bipolar I	Tipo
1	+	0	*C*	1	+	*C*
2	+	0	*C*	2	+	*C*
3	+	0	*C*	3	+	*C*
4	0	+	*L*	4	–	*L*
5	0	+	*L*	5	–	*L*
6	0	1	*L*	6	2	*L*

* + = indica cargas factoriales positivas; - = indica cargas factoriales negativas; 0 = cargas cero; *L* = escalas liberales; *C* = escalas conservadoras.

tanciales y negativas y los ceros indican cargas cercanas a cero. La teoría dualista (*A*) implica, por supuesto, dos factores ortogonales; y la teoría bipolar (*B*) implica un factor con cargas positivas y negativas sustanciales. La *A* y la *B* de la tabla expresan de manera sucinta los dos modelos implicados por las dos teorías. Si se graficaran las "cargas" de la teoría dualista se verían como las de la figura 34.6 del capítulo 34. Los puntos de las cargas de la teoría bipolar se grafican en un solo eje, con las cargas positivas en un extremo del eje y las cargas negativas en el otro extremo.

Como se mencionó con anterioridad y se reitera ahora, a los investigadores que utilizan el análisis estructural de covarianza les gusta desarrollar modelos en diagramas de

FIGURA 35.2 *Medidas observadas (escalas)* $x_1, x_2, \ldots, x_6$; ξ_1, ξ_2: *Xsi 1: conservadurismo (C); Xsi 2: liberalismo (L);* $\lambda_{11}, \lambda_{21}, \lambda_{31}, \ldots$: *lambdas, cargas factoriales;* $\delta_1, \delta_2, \ldots$: *delta 1, delta 2, ...: términos del error*



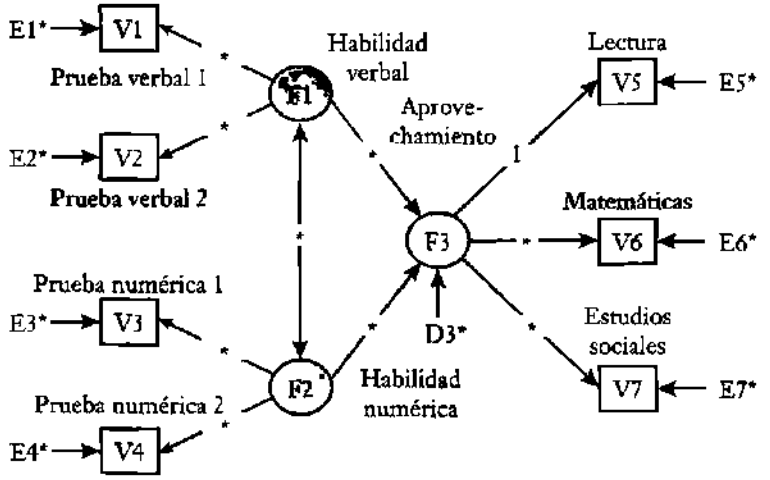
ruta. Los diagramas de ruta de los dos modelos factoriales se presentan en la figura 35.2. (Véase capítulo 33 para encontrar una explicación sobre los diagramas de ruta.) En este momento tan sólo se dibujará el modelo de ruta en términos de la notación del EQS. El principiante encontrará la notación del EQS mucho más fácil de entender que la del LISREL. El modelamiento del EQS sólo utiliza cuatro símbolos: V para variables medidas, F para variables latentes, E para representar el error de medición para las variables V, y D se utiliza para representar el error de medición para las variables latentes. El LISREL resulta más complicado.

A la luz de la naturaleza de este libro, es más congruente para el principiante emplear la notación de modelado dada por el EQS. El modelado del LISREL utiliza bastantes letras griegas para la designación de ciertos componentes o parámetros. Quizás ello asuste a algunos estudiantes y evite que aprendan una metodología de investigación y análisis sumamente importante.

En este ejemplo, $x_1, x_2, \ldots, x_6$ son las variables observadas; x_1, x_2 y x_3 son medidas del conservadurismo; y x_4, x_5 y x_6 son medidas del liberalismo, y se escribirían V1, V2, V3, V4, V5 y V6 en el EQS, donde V1, V2 y V3 miden *conservadurismo* y V4, V5 y V6 miden *liberalismo*. En ambas notaciones las variables observadas están indicadas por medio de cajas (por ejemplo, cuadrados o rectángulos); las variables no observadas latentes o factores, por medio de círculos o elipses. En el modelo dualista F_1 y F_2 se utilizan para representar *conservadurismo* y *liberalismo*. Los términos de error en el EQS se escriben E1, E2, E3, E4, E5 y E6. En el modelo de ruta presentado en la figura 35.3 se encuentra un asterisco junto a cada valor E, lo cual indica que los errores o valores únicos se estimarán por medio de los datos. La correlación entre las variables latentes también se va a estimar, por lo que también se especifica con un asterisco. Puesto que la teoría dice que el conservadurismo y el liberalismo son factores distintos y separados, se predice que $r_{12} = 0$.

El diagrama de la teoría bipolar es más fácil de explicar. Se tienen, evidentemente, las mismas seis x o variables observadas, y los mismos seis términos de error. También existen seis cargas factoriales; sólo existe un factor o F_1. En el modelo dual hay doce cargas factoriales, pero se predice que seis de ellas serán positivas y sustanciales, y el resto se fuerzan a ser iguales a cero. Los valores "forzados" o "fijos" se mantienen durante los

FIGURA 35.3 *Influencia de la habilidad sobre el aprovechamiento (ejemplo ficticio)*



cálculos. En el modelo de bipolaridad existen seis cargas factoriales: tres positivas (las rutas de las flechas están marcadas con +) y tres negativas (marcadas con –). En otras palabras, se predicen las dos matrices factoriales de la tabla 35.1, con excepción de que en la tabla sólo se usan signos + y – en lugar de cargas factoriales.

Para determinar cuál de los dos modelos está más cercano a la "realidad" empírica, es necesario probar cada uno de forma separada y después probar uno en contra del otro. Esto se hace utilizando la información o los datos que se tienen: las correlaciones entre las variables observadas x_1, x_2,..., x_6. Esta matriz de correlaciones, **R**, es una matriz de covarianza. Las variables (o escalas de actitud) 1, 2 y 3 son medidas del *conservadurismo*; y las variables 4, 5 y 6 son medidas del *liberalismo*. La hipótesis dualista predice correlaciones positivas y altas entre 1, 2 y 3; y correlaciones positivas y altas entre 4, 5 y 6. La hipótesis dualista también predice correlaciones de cero o cercanas a cero entre las variables de *C* (1, 2 y 3) y las variables de *L* (4, 5 y 6). Éstas se llaman *correlaciones cruzadas*.

El método que en realidad se utiliza en el análisis estructural de covarianza es el siguiente. Los datos se analizan de acuerdo con el modelo establecido; en este caso, el modelo dualista: dos factores ortogonales (figura 35.2A). A partir de los parámetros estimados por el análisis de los datos, análisis factorial en este caso, se calcula una matriz **R** utilizando los parámetros estimados del modelo teórico, lo cual se hace escribiendo ecuaciones para cada una de las V.

Para ayudar a comprender con mayor claridad lo que se hace y por qué, primero se establecieron las dos teorías en diagramas de ruta. El lector quizá piense que la siguiente explicación es redundante; pero los autores han encontrado que es útil para el aprendizaje y la comprensión de este método. Los investigadores del comportamiento que utilizan el "modelamiento" o "modelamiento causal", como se le llama, usan los diagramas de ruta para ayudar a conceptualizar los problemas de investigación que están estudiando y, casi más importante, para aprender las implicaciones empíricas de las teorías que se ponen a prueba. Resulta sumamente recomendable que los estudiantes traten de representar cualquier problema de investigación bajo estudio, en un diagrama de ruta, pues éste obliga a conceptualizar y obtener las estructuras básicas de los problemas. En cualquier caso, las "teorías" dualista y bipolar sobre las actitudes sociales se han establecido en los dos diagramas de ruta *A* y *B* de la figura 35.3. En dichos diagramas de ruta se acostumbra utilizar cuadrados para las variables observadas y círculos para las variables no observadas o latentes. Las flechas unidireccionales sirven para indicar influencias y las flechas bidireccionales para indicar correlaciones. Por ejemplo, si se realizara un análisis factorial de factores principales del presente problema, se estimarían 12 cargas factoriales: a_{11}, a_{12}, a_{21}, a_{22},..., a_{61}, a_{62}. Sin embargo, el problema en el EQS o en el marco conceptual de las estructuras de covarianza resulta diferente, debido a que ya se ha especificado que seis cargas factoriales o *a* se estimarán; las cargas restantes se fuerzan a ser iguales a cero en la hipótesis dualista. Para asegurarse de que se percibe y comprende la diferencia, se establecieron las dos matrices factoriales en la tabla 35.2. Note que hay 12 cargas factoriales a calcularse en A, con un análisis factorial común, y sólo seis cargas a calcularse en B, la solución de estructuras de covarianza forzada por los ceros, debido a la hipótesis dualista.

La diferencia entre los dos modelos es sorprendente —y muy importante—. En el análisis factorial común se estiman todas las cargas factoriales; pero en el análisis estructural de covarianza sólo se estiman aquellas cargas factoriales relacionadas con las hipótesis. El resto se fuerzan a ser iguales a cero —una estructura simple perfecta—. Para enfatizar los puntos establecidos, los parámetros estimados reales se muestran en la tabla 35.4. Los factores finales rotados de un análisis factorial común se presentan en *a*), y la solución forzada de las estructuras de covarianza se muestra en *b*). Se podría plantear la pregunta: ¿qué les sucede a las cargas factoriales donde se encuentran los ceros en *b*)? El punto es

TABLA 35.2 *Matrices factoriales del a)análisis factorial común y del b)análisis factorial forzado del EQS*[a]

	a). Análisis factorial ordinario			*b*). Análisis factorial forzado del EQS		
Variables	I	II	Variables	I	II	Tipo[a]
1	a_{11}	a_{12}	1	a_{11}	0	*C*
2	a_{21}	a_{22}	2	a_{21}	0	*C*
3	a_{31}	a_{32}	3	a_{31}	0	*C*
4	a_{41}	a_{42}	4	0	a_{42}	*L*
5	a_{51}	a_{52}	5	0	a_{52}	*L*
6	a_{61}	a_{62}	6	0	a_{62}	*L*

[a] *C* = conservador, *L* = liberal.

que *b*) expresa la forma "pura" de la hipótesis dualista. Como se dijo anteriormente, la computadora está programada para realizar los cálculos manteniendo intactos los ceros de la tabla 35.2 y de la tabla 35.3. ¿Pero qué pasa con las cargas bastante grandes y negativas, –.44 y –.36 en *a*), el análisis factorial convencional? Ambas son altas, negativas y estadísticamente significativas, en oposición a la hipótesis dualista. Son desviaciones a partir del modelo dualista. Entonces, la pregunta clave es: ¿las desviaciones son lo suficientemente grandes para invalidar la hipótesis la cual incluye ceros? Se retomará este punto en breve.

El modelo de la figura 35.2A requiere cálculos de los términos de error, E. Se calcularon los seis términos de error; aunque aquí no interesa el método de los cálculos. La estimación de las varianzas y covarianzas es mucho más interesante y relevante para la hipótesis dualista, ya que ésta expresa las relaciones entre los factores F1 y F2. Recuerde que la hipótesis dualista incluía la correlación entre los dos factores: sería de cero o cercana a cero. Observe nuevamente la figura 35.2 y note que, de acuerdo a la hipótesis dualista, $r_{12} = 0$. Aunque r_{12} puede forzarse para que sea igual a cero, en su lugar se eligió permitir al EQS estimar la correlación entre los factores, por razones que se explicarán más adelante. Para reflejar esto en la figura 35.2A, se cambiaría r_{12} por un asterisco. Se establece que las varianzas de F1 y F2 sean iguales a 1.00, se estiman las varianzas de E1, E2, ..., E6, y se

TABLA 35.3 *Matrices factoriales obtenidas: a) convencional (con rotación) y b) factores forzados del EQS*[a]

	a) Factores convencionales			*b*) Factores forzados del EQS		
Variables	I	II	Variables	I	II	Tipo
1	.69	.19	1	.65	0	*C*
2	.70	.33	2	.87	0	*C*
3	.68	.14	3	.63	0	*C*
4	–.44	.51	4	0	.71	*L*
5	–.05	.64	5	0	.54	*L*
6	–.36	.55	6	0	.63	*L*

[a] El análisis factorial convencional fue el método de factores principales con rotación varimax; el método de EQS fue el de probabilidad máxima. Todas las cargas de *b*) son estadísticamente significativas. La correlación entre los dos factores fue de –.15, que no resultó estadísticamente significativa.

especifica r_{12} como "libre". (Recuerde que cuando un parámetro es "libre", el programa estima su valor.)

En el análisis, $r_{12} = -.157$ no es estadísticamente significativa. Por lo tanto, en efecto, los dos factores son ortogonales, lo cual es consistente con la hipótesis dualista. Tenga en cuenta que la teoría dice que el *conservadurismo* y el *liberalismo* son dimensiones separadas e independientes de las actitudes sociales, lo que quiere decir, evidentemente, que la correlación entre ellas es cero (o cercana a cero).

Sin embargo, la pregunta crucial es: ¿el modelo completo es congruente con los datos? El modelo completo de la hipótesis dualista está expresado en la figura 35.2A. Siguiendo las reglas del EQS, se instruye a la computadora para que estime las seis cargas factoriales, a_{11}, a_{21}, a_{31}, a_{42}, a_{52}, a_{62}; mientras mantiene las limitaciones a cero en la matriz. Además se especifica que se calculen los términos de error de las seis ecuaciones. Además se deben especificar cuáles serán las relaciones entre los dos factores. Por lo tanto, se le debe indicar al EQS qué hacer con las varianzas de los factores F1 y F2; ello se logra al instruir al EQS para que estime r_{12}, la correlación entre F1 y F2. En el análisis factorial tradicional, lo anterior implicaría estimar los valores en la matriz Φ (véase capítulo 34). Siguiendo un procedimiento iterativo, el programa computacional estima los 13 valores que se especificó deben estimarse, utilizando las correlaciones entre las seis variables como datos de entrada (tabla 35.2). También fuerza los ceros de la tabla 35.2 y establece que las varianzas de las variables latentes (o factores) sean iguales a 1.00. Las cargas factoriales se presentan en la tabla 35.3B, y $r_{12} = -.157$. Los seis términos de error son .76, .50, .78, .71, .84 y .77. ¿Estos valores son congruentes con los datos o, de manera alterna, se "ajusta" el modelo dualista con los datos? Antes de responder las preguntas es necesario mencionar que existe una cantidad de otros puntos metodológicos importantes que no se exponen aquí, como los supuestos que subyacen al análisis. Uno de los supuestos es que la distribución de las variables observadas o medidas es normal. Otro supuesto o requisito es la identificación: el problema de la estructura de la covarianza debe establecerse de tal manera que todos los parámetros estimados puedan identificarse. Existen problemas de investigación donde tales supuestos puedan no cumplirse. La idea central detrás de la evaluación de la "bondad de ajuste" de un modelo teórico es simple y poderosa. Emplee los valores de los parámetros estimados y los valores forzados para calcular una matriz de correlación predicha o reproducida o ajustada, $\mathbf{R}^*$. En este caso la matriz $\mathbf{R}^*$ puede generarse multiplicando los renglones de la tabla 35.3: $r_{12}^* = (.65)(.87) + (0)(0) = .57$; $r_{13} = (.65)(.63) + (0)(0) = .41$; $r_{23} = (.87)(.63) + (0)(0) = .55$, etcétera. Entonces esta $\mathbf{R}^*$ se compara con la matriz de correlación obtenida u observada, $\mathbf{R}$, lo cual puede realizarse restando $\mathbf{R}^*$ de $\mathbf{R}$, o $\mathbf{R} - \mathbf{R}^*$. Dicha matriz de diferencias se llama una *matriz residual*. En el análisis estructural de covarianza los residuales casi siempre se analizan con uno de tres modelos de funciones de ajuste:

1. Mínimos cuadrados sin ponderar
2. Mínimos cuadrados generalizados o ponderados
3. Máxima verosimilitud

Como se estudió en el capítulo 34, existe un número de estadísticos diferentes de bondad de ajuste. El más antiguo de ellos —la chi cuadrada— algunas veces se informa, pero no es utilizado como el único estadístico, debido a que su evaluación se basa en el tamaño de la muestra. Conforme el tamaño de la muestra se agranda, pequeñas diferencias se vuelven estadísticamente significativas, e indican una falta de ajuste. Bentler (1980) ofrece una excelente revisión de los estadísticos de bondad de ajuste y sugiere el uso de estadísticos que no dependan del tamaño de la muestra. El programa EQS, desarrollado por Bentler (1986), originalmente utilizaba un estadístico de ajuste de este tipo, llamado

índice de ajuste normado de Bentler-Bonett o IAN, el cual ahora es obsoleto. El índice de ajuste actual de elección es el índice de ajuste comparativo (IAC), y un valor de .95 o mayor es representativo de un buen ajuste entre el modelo y los datos. Los valores del IAC menores de .95 indican al investigador que existe posibilidad de mejorar la manera en que se especifica el modelo. En esencia dice que el modelo no se ajusta muy bien a los datos. Si se obtienen valores alrededor de .95 o mayores, el ajuste de los datos al modelo es bastante bueno y es poco probable que cualquier reespecificación posterior del modelo altere mucho el índice. El LISREL, desarrollado por Joreskog y Sorbom (1993), originalmente utilizaba un estadístico de bondad de ajuste diferente, llamado la Raíz del Cuadrado Medio Residual (RMR por sus siglas en inglés). No obstante, ahora todos los programas populares poseen los estadísticos de ajuste más comunes, tales como el índice de bondad de ajuste y el índice ajustado de bondad de ajuste (véase Comrey y Lee, 1992; o Keith, 1999, para encontrar mayores explicaciones respecto a tales índices).

A partir de los resultados del EQS, el estadístico chi cuadrada para el modelo dualista es 121.253, basado en 8 grados de libertad. El valor de probabilidad para el estadístico chi cuadrada es menor que .001, lo cual indica que los datos no se ajustan al modelo. El índice de ajuste normado Bentler-Bonett fue de .840, lo cual indica que es posible hacer mejoras para lograr un mejor ajuste.

Ahora se revisarán las ideas que subyacen al método. El principio es: *a menor tamaño de los residuales, mejor será el ajuste; a mayor tamaño de los residuales, más pobre será el ajuste.* Si la hipótesis o modelo es válido empíricamente, menos diferencia habrá entre la matriz de covarianza (correlación) generada a partir del modelo, $\mathbf{R}^*$, y la matriz de correlación observada, **R**. Ambas situaciones se reflejan en la matriz de residuales, $\mathbf{R} - \mathbf{R}^*$, y en medidas, como el IAN, que refleja la magnitud de los residuales. Nuevamente, a mayor tamaño de los residuales, más pobre será el ajuste. (Los tres programas computacionales de estructura de covarianza imprimen la matriz residual de manera obligatoria.)

Las implicaciones empíricas de la hipótesis de bipolaridad se describen en la figura 35.2B. Evidentemente existe sólo un factor: F1. Las medidas del conservadurismo V1, V2 y V3 (o x_1, x_2 y x_3) están marcadas con "+"; y las de V4, V5 y V6 (o x_4, x_5 y x_6) están marcadas con "–", lo cual es consistente con la hipótesis de bipolaridad. Es decir, se espera un factor bipolar donde las medidas *conservadoras* tengan signos positivos; y las medidas *liberales*, signos negativos (o a la inversa). Las seis cargas factoriales estimadas por el EQS sobre *un* factor fueron .67, .83, .65, –.25, .12 y –.15. $\chi^2 = 313.143$, basado en 9 grados de libertad. El valor de probabilidad para el estadístico chi cuadrada es menor a .001. El IAN Bentler-Bonett = .586. Tales valores indican que la bondad de ajuste del modelo bipolar fue mucho peor que la del modelo dualista.

Las cargas factoriales son interesantes e informativas. Las de las tres medidas del *conservadurismo*, V1, V2 y V3, son positivas y sustanciales; las de las medidas del *liberalismo* son todas bajas. Evidentemente el modelo de un factor resulta inadecuado: las tres medidas liberales se "pierden". La χ^2 también es significativa, lo que indica una falta de ajuste. Ahora observe los residuales en la mitad superior de la tabla 35.5. Note cuidadosamente que los residuales de r_{45}, r_{46} y r_{56} son sustanciales: .416, .393 y .389. Las correlaciones entre las medidas liberales, V4, V5 y V6 se "perdieron" con la solución de un solo factor, el modelo para la hipótesis bipolar. Parece ser que el modelo de bipolaridad no ha sido muy exitoso. Por otro lado, el modelo dualista se desempeñó mejor en todos los cálculos.

Ahora se realiza una prueba final: se comparan directamente los dos modelos, lo cual se hace por medio de las pruebas de χ^2. La χ^2 para el modelo bipolar fue de 313.143, con nueve grados de libertad, mientras que la χ^2 para el modelo dualista fue de 121.253, con 8 grados de libertad. Recuerde que antes se pidió a la computadora que estimara el valor de r_{12}, aunque estrictamente hablando, se debió haber fijado en cero, o $r_{21} = 0$. Ello se debe a

que el modelo dualista puro predice factores ortogonales. Una de las razones principales para hacer esto fue "agotar" un grado de libertad para comparar las χ^2 de los dos modelos. La prueba directa es $\chi^2 + 2_{bip} - x_{du} - 121.253 = 191.89$. También se restan los grados de libertad: 9 – 8 = 1. Si no se hubiera estimado r_{12}, los grados de libertad para los dos modelos hubieran sido los mismos, haciendo imposible una comparación de χ^2. Se evalúa χ^2 = 191.89, con gl = 1; es altamente significativa, lo cual indica la superioridad de la hipótesis dualista (puesto que la χ^2 del modelo bipolar es significativamente mayor que la χ^2 del modelo dualista). Si no hubiera una diferencia significativa entre las χ^2 de los dos modelos, entonces la hipótesis bipolar sería tan "buena" (o tan "pobre") como la hipótesis dualista. Por consiguiente, no es posible inferir que una hipótesis sea más satisfactoria que la otra. Recuerde que un modelo que es congruente con los datos tendrá una χ^2 no significativa estadísticamente. Sin embargo, si la diferencia entre las χ^2 es significativa, entonces se infiere que el modelo con el valor χ^2 mayor resulta *menos* satisfactorio que el modelo con el valor χ^2 menor. Otra forma de explicarlo es que si la diferencia entre las χ^2 es significativa, prueba la importancia de los parámetros que diferencian a los modelos.

Lo arriba expuesto es difícil de mostrar y explicar, de acuerdo con la manera en que se ha realizado el problema. Un enfoque más elegante es el siguiente. Se establece el modelo dualista de la forma en que se hizo antes. Después se establece el modelo bipolar exactamente de la misma forma, con excepción del término r_{12}. Para el modelo dualista se estima r_{12} igual que antes. Esto producirá una χ^2 con gl = 8. Ahora se establece el modelo bipolar fijando r_{12} = 1.00, con gl = 9, lo cual producirá exactamente los mismos estimados de los parámetros que si se le hubiera indicado al programa que había sólo un factor, excepto que las cargas del único factor aparecerán en dos factores. Puesto que la correlación entre los dos factores es 1.00, el efecto neto es el mismo que con un factor. La prueba de las hipótesis alternativas, $\chi^2_{bip} - \chi^2_{dual}$ será igual a la anterior. No obstante, ahora queda claro que los dos modelos difieren tan sólo en un parámetro: Φ_{21}. Ésta es una de las razones por las que se calcula Φ_{21} o r_{12} en el modelo dualista: para hacer una prueba de hipótesis alternativas debe existir una diferencia en los grados de libertad. Además, un modelo debe ser un subconjunto del otro modelo, lo que significa que ambos modelos estiman los mismos parámetros, con excepción (en este caso) de un parámetro.

Influencias de las variables latentes: el sistema EQS completo

En el ejemplo anterior sobre las actitudes se utilizó sólo una parte de la estructura de covarianza o del sistema de modelamiento de ecuaciones estructurales. Si la solución que se pretendía era el análisis factorial ordinario de primer orden, entonces lo que se hizo era todo lo necesario. Sin embargo, los problemas más interesantes estudian las relaciones entre variables independientes y variables dependientes. Antes de explicar las propiedades formales del sistema, se examinará un ejemplo ficticio simple. Se estableció el diagrama de ruta del ejemplo, para tener algo concreto a qué referirse; dicho diagrama se muestra en la figura 35.3. El ejemplo es un pequeño modelo de *habilidad* y *aprovechamiento*. En efecto, se dice: la *habilidad verbal* y la *habilidad numérica* influyen en el aprovechamiento de forma positiva. Aunque tal vez el ejemplo no sea muy interesante, posee la virtud de ser obvio y fácil de comprender. Aquí no se intenta probar hipótesis alternativas, aun cuando existe una diversidad de posibilidades. Tan sólo se busca transmitir la esencia del sistema.

Observe este sistema y su función desde el punto de vista de la regresión. Primero, considere la parte izquierda de la figura 35.3. Hay cuatro pruebas: *prueba verbal 1*, *prueba*

verbal 2, *prueba numérica 1* y *prueba numérica 2*, V1, V2, V3 y V4 usando notación del EQS. Son variables dependientes medidas. Se calcula la matriz de correlación 4×4, se analiza factorialmente y se obtienen dos factores, F1 y F2, como en la figura 35.3. Las flechas con asteriscos que señalan hacia las variables dependientes, a partir de F1 y F2, contienen las cargas factoriales: a_{11}, a_{21}, a_{32} y a_{42}. Las otras cargas se fijan en cero, tal como se hizo antes con la hipótesis dualista de actitudes:

Pruebas	I	II
1	a_{11}	0
2	a_{21}	0
3	0	a_{32}
4	0	a_{42}

Se pueden considerar las a como coeficientes de regresión. La ecuación de regresión para V1 es:

$$V1 = a_{11}F1 + E1$$

Se busca la regresión de V1 a partir de F1, tal y como se buscó la regresión de y a partir de x, o de y a partir de $x_1, x_2, \ldots$ Es factible pensar que las cargas factoriales a, tienen la misma función que los coeficientes de regresión, b o β, del capítulo 32. El mismo razonamiento se aplica a la parte derecha de la figura 35.3: se escribe la regresión de V5 a partir de F3 como:

$$V5 = a_1F3 + E5$$

En dicha estructura de covarianza existen dos análisis factoriales o sistemas de regresión separados: uno en la parte izquierda, y otro en la parte derecha. Cualquiera de las dos partes puede utilizarse para el análisis factorial confirmatorio o de comprobación de hipótesis, como cuando se probaron las hipótesis dualista y bipolar. No obstante, lo que es más interesante e innovador es plantear y responder preguntas de investigación acerca de la regresión de la variable latente a partir de otra(s) variable(s) latente(s). En efecto, se pregunta sobre las relaciones entre F1, F2 y F3, o la regresión de F3 a partir de F1 y F2, considerando a F1 y F2 como variables independientes, y a F3 como variables dependientes. Esto es lo que hace la ecuación estructural o el análisis estructural de covarianza.

El problema de investigación de la figura 35.3 se expresa como la relación multivariada entre las variables independientes y las variables dependientes. Puede resolverse, por ejemplo, utilizando la correlación canónica, la cual expresaría la relación general entre las V del lado izquierdo (por ejemplo, V1, V2, V3 y V4) y las V del lado derecho (V5, V6 y V7). Pero la correlación canónica no es capaz de refinar las relaciones. Consigue la relación entre dos conjuntos de variables utilizando *todas* las variables. Por lo general, no tiene nada que ver con las variables latentes. El modelo e hipótesis implicadas en la figura 35.3 indican, en efecto, que las variables dependientes medidas de la parte izquierda reflejan dos factores: *habilidad verbal* (V1 = *prueba verbal* 1, y V2 = *prueba verbal* 2) y *habilidad numérica* (V3 = *prueba numérica* 1, y V4 = *prueba numérica* 2). Las variables latentes son *habilidad verbal*, F1 y *habilidad numérica*, F2. Las tres pruebas de aprovechamiento son *lectura*, V5; *matemáticas*, V6; y *estudios sociales*, V7. Se presume que miden un factor, aprovechamiento —en otras palabras, una hipótesis unifactorial—. Note cuidadosamente que las hipótesis que no son satisfactorias, en el sentido de que no son congruentes con los datos, pueden establecerse fácilmente, invalidando al modelo completo. Por ejemplo, las variables V5, V6 y V7, que se dijo medían el reflejo de *un* factor o variable latente, podrían

ser incorrectas. Quizá se necesiten dos factores. Es decir, la figura 35.3 tiene un factor, F3, para el *aprovechamiento*, pero pueden ser en realidad dos factores, F3 y F4. Después de todo, V5 es una prueba de lectura y V6 es una prueba de matemáticas, y se sabe que, por lo general, estos dos factores son diferentes. Si esto es así, entonces el modelo de la figura 35.3 es deficiente al respecto.

Establecimiento de la estructura del EQS

Por último llega la relación crucial: la de F1 y F2, las variables independientes latentes; y F3, la variable dependiente latente. La hipótesis sustantiva establecería qué tanto la *habilidad verbal* F1, y la *habilidad numérica* F2 influyen en el *aprovechamiento* F3.

Dicha hipótesis no es demasiado fascinante, más bien resulta sensible para ejemplificar y explicar. Para probarla, se deben plantear el problema y modelo de la figura 35.3 en proposiciones y estructura del programa EQS. Se trata de un paso complicado y crucial en el EQS. Debido al riesgo de provocar tedio, es menester seguir las ideas y edificarlas en ecuaciones y ecuaciones de matrices, después de descubrir las ecuaciones de una variable individual. Primero, las ecuaciones para el lado izquierdo de la figura 35.3:

$$\begin{aligned} V1 &= .3*F1 + E1 \\ V2 &= .3*F1 + E2 \\ V3 &= .3*F2 + E3 \\ V4 &= .3*F2 + E4 \end{aligned} \tag{35.2}$$

".3*" indica que existen coeficientes que se estimarán a partir de los datos; éstos son las cargas factoriales. El valor ".3" es un valor inicial arbitrario de "adivinación" para la estimación.

Para la especificación del EQS, se escriben las mismas ecuaciones en forma de matriz:

$$\begin{bmatrix} V_1 \\ V_2 \\ V_3 \\ V_4 \end{bmatrix} = \begin{bmatrix} a_{11} & 0 \\ a_{21} & 0 \\ 0 & a_{32} \\ 0 & a_{42} \end{bmatrix} \begin{bmatrix} F_1 \\ F_2 \end{bmatrix} + \begin{bmatrix} E_1 \\ E_2 \\ E_3 \\ E_4 \end{bmatrix} \tag{35.3}$$

Donde las *a* se estiman a partir de los datos. (El lector debe hacer una pausa aquí, estudiar la figura 35.4 y las ecuaciones 35.2 y 35.3, e intentar comprender su significado.)

El lado derecho resulta un poco más fácil:

$$\begin{aligned} V5 &= 3*F3 + E5 \\ V6 &= 3*F3 + E6 \\ V7 &= 3*F3 + E7 \end{aligned} \tag{35.4}$$

En forma de matriz:

$$\begin{bmatrix} V_5 \\ V_6 \\ V_7 \end{bmatrix} = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} F_3 + \begin{bmatrix} E_5 \\ E_6 \\ E_7 \end{bmatrix} \tag{35.5}$$

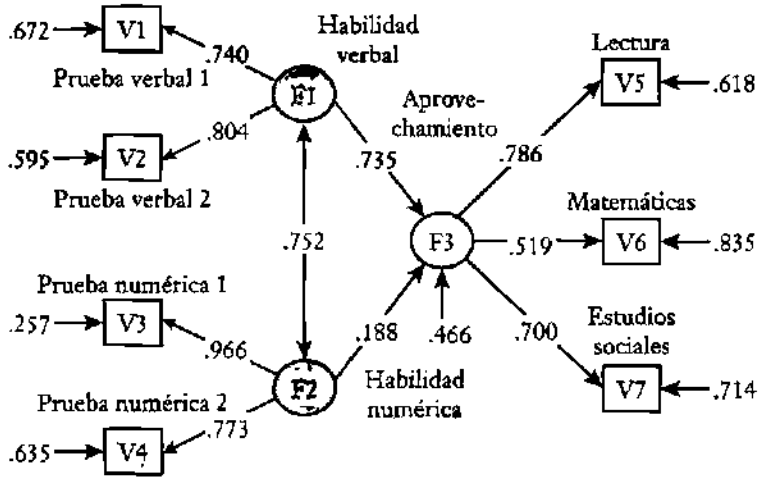
Se sintetizó una matriz de correlación ficticia, de tal manera que la solución del EQS apoye el modelo del diagrama de ruta de la figura 35.3, y las ecuaciones que se escribieron

con base en el diagrama. Los resultados fueron satisfactorios. El estadístico chi cuadrada fue estadísticamente significativo, lo cual indica una posible falta de ajuste. Sin embargo, como se mencionó anteriormente en este capítulo, existen otros índices que quizá sean mejores indicadores del ajuste. El *índice de ajuste normado Bentler-Bonett* (IAN) fue de .94, lo cual constituye un muy buen valor. Por lo común, se considera que cualquier valor de .90 o mayor indica un modelo bien ajustado. Otros índices calculados por el EQS, como el índice de ajuste comparativo (IAC = .95), apoyan la conclusión de que el modelo resultó satisfactorio.

Aunque los parámetros de los análisis factoriales (o análisis de regresión) para las partes izquierda y derecha de la figura 35.3 también son satisfactorios, no se informan. Lo anterior se debe a que el interés aquí es la comprobación del modelo respecto a su congruencia con los datos, en este caso una matriz de correlación. Además, también existe el interés en evaluar las relaciones entre las variables latentes *habilidad verbal* y *habilidad numérica*, por una parte, y el *aprovechamiento*, por la otra. Los valores que expresan estas influencias se presentan en la figura 35.4, la cual es igual a la figura 35.3, con excepción de que se muestran los parámetros estimados.

Anteriormente se dijo que el análisis estructural de covarianza y el programa computacional utilizado para realizar los cálculos complejos necesarios no eran sencillos de aprender. Aun utilizando el modelo más simple del EQS, quizá resulte difícil para el inexperto. Los 10 puntos antes mencionados son importantes para quienes deseen utilizar dicho método extremadamente poderoso. Sin embargo, incluso éstos no son fáciles de comprender. Entonces, ¿para qué molestarse aprendiéndolos? ¿No es posible realizar los análisis factoriales y los análisis de regresión de manera separada, para que el investigador del comportamiento se complique menos? Sí y no. Los análisis factoriales separados de las variables del lado derecho e izquierdo de la figura 35.3 pueden, de hecho, realizarse de forma separada. En efecto, los estudios psicométricos y analíticos factoriales deben realizarse antes de utilizar el EQS o el modelado de ecuación estructural. Pero obviamente el análisis de regresión apenas descrito no puede llevarse a cabo con problemas de investigación complejos que incluyan variables latentes y medidas indirectas y directas. De hecho,

FIGURA 35.4 *Mismo diagrama de ruta de la figura 35.3, con estimados paramétricos*



es posible intentar diversos enfoques para el análisis de los datos. Pero parece no haber una forma simple para estudiar conjuntos de relaciones complejas y para probar la congruencia de modelos teóricos con datos observados. Las ideas del análisis estructural de covarianza son matemática y estadísticamente poderosas, conceptualmente penetrantes y estéticamente satisfactorias. La creación del EQS, LISREL, AMOS y otros programas computacionales similares son logros bastante ingeniosos, productivos y creativos. Constituyen, hasta el momento, el más alto desarrollo del pensamiento analítico y científico del comportamiento; es un desarrollo que une la teoría psicológica y sociológica con el análisis matemático y estadístico multivariado en una síntesis única y poderosa que quizá revolucionará la investigación del comportamiento. Es en este sentido que se dice que el análisis estructural de covarianza es la culminación de la metodología contemporánea.

Estudios de investigación

En el relativamente poco tiempo que el análisis estructural de covarianza y los programas computacionales para realizarlo han funcionado y estado disponibles —desde inicios y mediados de los años setenta— el método se ha utilizado de manera fructífera en diversos campos. Algunos de dichos estudios son reanálisis de datos existentes; otros son estudios que se concibieron teniendo en mente un análisis estructural de covarianza (véase sugerencia de estudio 2). El primer estudio sobre la estructura de las actitudes, estudiado en el presente capítulo, constituye sólo uno de los 12 conjuntos de datos sobre actitudes que se reanalizaron con modelos de ecuación estructural o análisis estructural de covarianza. La mayor parte de la evidencia apoyó la hipótesis dualista (véase Kerlinger, 1980). Joreskog y sus colaboradores reanalizaron los datos de diversos estudios psicológicos y sociológicos (véase Magidson, 1979). El primer estudio descrito a detalle a continuación consiste en un reanálisis estructural de covarianza de los datos de un estudio grande sobre participación política en Estados Unidos.

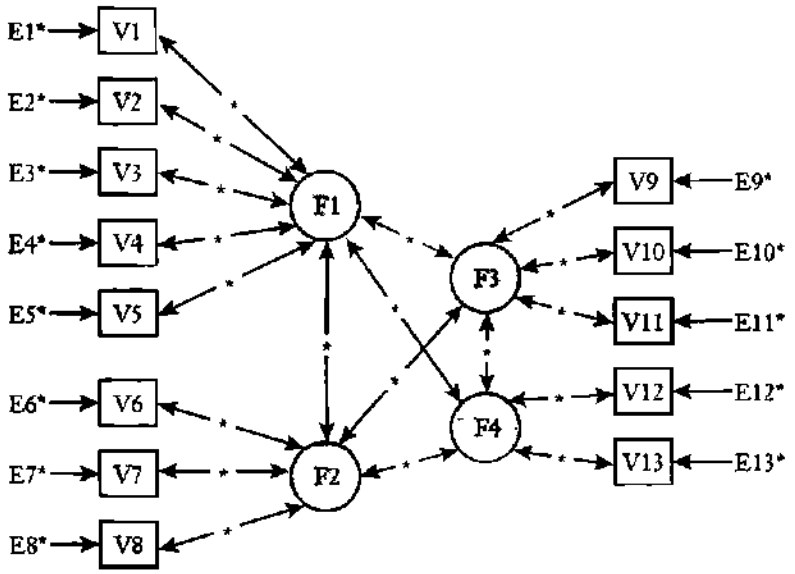
Bentler y Woodward (1978) utilizaron el análisis estructural de covarianza para reanalizar datos del Head Start —con resultados deprimentes—. Encontraron que el programa Head Start no tenía efectos significativos sobre las habilidades cognitivas de los niños que participaron en el programa. Judd y Millburn (1980) estudiaron la estructura de la actitud del público general en Estados Unidos. Utilizaron datos de panel de encuestas realizadas en 1972, 1974 y 1976, investigaron la opinión de Campbell, Converse, Miller y Stokes (1960) de que el público general no posee actitudes sociales estables y significativas. Encontraron que el público sin educación *sí* posee predisposiciones ideológicas consistentes.

Verba y Nie: participación política en Estados Unidos

En un estudio sobre participación política, Verba y Nie (1972) consideraron, a partir de la teoría política, que deberían existir cuatro factores detrás de 13 variables de participación política. Dichos factores y variables se presentan en la figura 35.5. Su estudio consistió en un análisis factorial confirmatorio. Parecían estar correctos en su hipótesis estructural, y se aplaudió su cuidadoso y competente trabajo. Pero el análisis factorial ha sido criticado, entre otras cosas, por su falta de rigor. ¿La hipótesis estructural de Verba y Nie puede someterse a una prueba más rigurosa? Permítase el uso del EQS en el problema. Sin embargo, note que Verba y Nie no utilizaron ecuaciones estructurales. Por consiguiente, los resultados aquí presentados provienen del reanálisis de sus datos.

El modelo de diagrama analítico de ruta que surge a partir de la discusión teórica de Verba y Nie se presenta en la figura 35.5. V1, V2, V3,..., V13 son las variables dependien-

FIGURA 35.5 *Diagrama de ruta (estudio de Verba y Nie)*



[a] Persuadir a otros sobre cómo votar; 2. Trabajo activo por un partido o candidato; 3. Acudir a junta o debate político; 4. Aportar dinero a un partido o candidato; 5. Participación en clubes políticos; 6. Votó en la elección presidencial de 1964; 7. Votó en la elección presidencial de 1960; 8. Frecuencia de votación en elecciones locales; 9. Trabajar con otros en problemas locales; 10. Formar un grupo para trabajar en problemas locales; 11. Participación activa en organizaciones comunitarias que resuelven problemas; 12. Contactar funcionarios locales; 13. Contactar funcionarios estatales y nacionales.

[b] Factor I: actividades de campaña (variables 1-5); factor II: votación (variables 6, 7 y 8); factor III: actividad de cooperación (variables 9-11); factor IV: contactar (variables 12 y 13).

tes medidas. Los componentes de error, asociados con cada variable dependiente, son E1, E2, E3,..., E13. Existen cuatro factores hipotetizados: F1, F2, F3 y F4; éstas son las variables independientes latentes. Se hipotetiza que los cuatro factores están correlacionados entre sí. Recuerde que cuando los factores están correlacionados, la solución es oblicua. Verba y Nie encontraron los siguientes factores: A-actividad de campaña (variables 1-5); B-votación (variables 6, 7 y 8); C-actividad de cooperación (variables 9-11); D-contactación (variables 12 y 13).

Se dieron instrucciones al EQS para calcular los estimados de los parámetros de la figura 35.5, y después de utilizar los parámetros para calcular una matriz de correlación predicha $\mathbf{R}^*$. Finalmente, para evaluar la adecuación del ajuste del modelo, de los cuatro factores oblicuos de la figura 35.5, se calcularon $\mathbf{R} - \mathbf{R}^*$, las diferencias o residuales y diversos estadísticos de "ajuste".

Los resultados generales apoyan el modelo de Verba y Nie de los cuatro factores oblicuos, a pesar de que la $\chi^2 = 406.648$, con $gl = 59$, es altamente significativa. El alto valor de χ^2 se debe claramente a la enorme N de 3 000 y, por lo tanto, no es una buena medida del ajuste. (Una solución idéntica con una N reducida de 300 produjo una $\chi^2 = 40.54$, que no es significativa.) La raíz del cuadrado medio residual (RCM) fue de .03. Este pequeño índice tan sólo reflejó los generalmente pequeños residuos. El índice de ajuste normado (IAN) Bentler-Bonett calculado por el EQS fue de .961, el cual es muy alto. El índice de

ajuste comparativo (IAC) fue muy alto, .966. Tales índices indican un muy buen ajuste. En síntesis, el ajuste del modelo de la figura 35.5 es bueno. El razonamiento teórico y el procedimiento de medición de Verba y Nie parecen ser adecuados. Ellos contribuyeron de forma significativa a la comprensión del proceso político y a la naturaleza y significado de la participación en el proceso político.

Brecht, Dracup, Moser y Riegel: relación entre la calidad marital y el ajuste psicosocial

En el capítulo anterior se expuso la investigación no experimental, que incluye aquellos estudios sin manipulación de las variables independientes. Por lo general, se estudia la variación entre las variables existentes y, en algunos casos, se implica débilmente una inferencia causal. Dichos estudios prevalecen en la investigación realizada en escenarios aplicados. Los investigadores que realizan investigación en escenarios aplicados por lo general no se dan el lujo de la asignación aleatoria o selección aleatoria. Keith (1999) afirma específicamente que una gran cantidad de estudios de investigación realizados en psicología escolar son de naturaleza no experimental. Otra área donde la investigación no experimental constituye la metodología dominante son las ciencias de la salud, en especial la investigación sobre el cuidado de pacientes. Las bases de datos sobre el cuidado de pacientes son grandes pero complejas y no experimentales. No obstante, es posible obtener información clave a partir de tales datos e investigación.

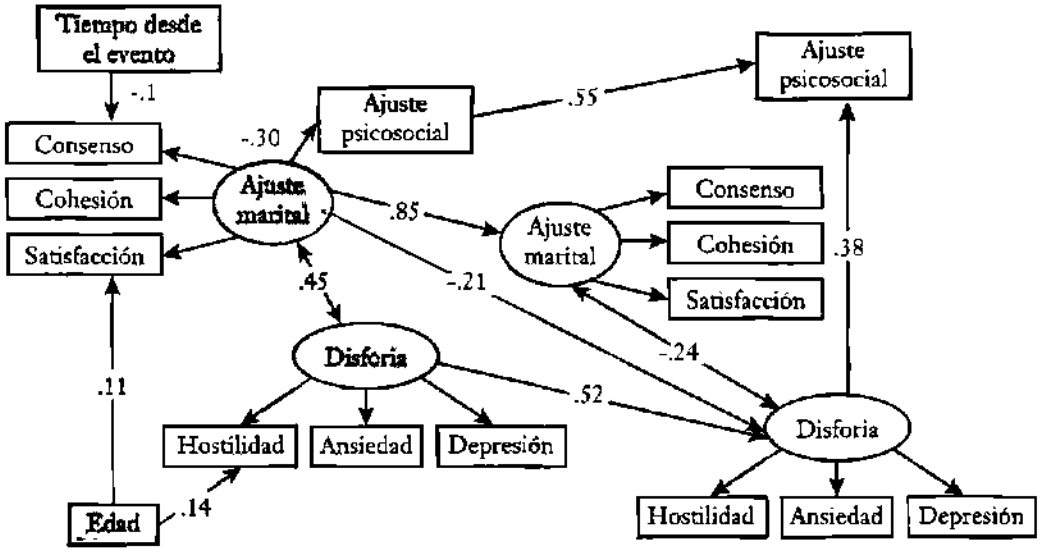
Un método fructífero para analizar los datos no experimentales sobre el cuidado de pacientes es el análisis estructural de covarianza o modelamiento de la ecuación estructural (MEE). Un estudio de este tipo que utilizó apropiada y exitosamente dicho método y, como tal demostró su valor, es el realizado por Brecht, Dracup, Moser y Riegel (1994).

Aquí los investigadores estudiaron el ajuste psicosocial de pacientes con enfermedad cardiaca. Investigación anterior sobre este tema sugería relaciones entre ciertas variables y el ajuste psicosocial, aunque no se ha definido la naturaleza precisa de la relación de las variables y el ajuste. Brecht *et al.* intentaron definir esto al señalar el hallazgo de que algunos pacientes se recuperan más rápido de la cirugía cardiaca que otros, así como de que también experimentan menos tensión emocional.

Bretch *et al.* plantearon la hipótesis de que la calidad de la relación marital, la disforia (ansiedad, depresión y hostilidad), la edad y el tiempo transcurrido desde la cirugía tenían posibles efectos directos e indirectos sobre el ajuste psicosocial. La muestra consistió en 198 pacientes cardiacos masculinos. Se tomaron mediciones de las variables en dos momentos diferentes. La primera fue al inicio del estudio (*momento 1*) y la segunda (*momento 2*) se realizó tres meses más tarde. La variable dependiente primaria era el *ajuste psicosocial*, el cual se midió utilizando la escala de ajuste psicosocial a la enfermedad (Psychosocial Adjustment to Illness Scale, PAIS). Calificaciones altas en la escala indicaban un mal ajuste. La calidad de la relación marital se midió utilizando la escala Spanier de ajuste diádico (Spanier Dyadic Adjustment Scale), y la disforia se midió a través del listado de adjetivos afectivos múltiples (Multiple Affect Adjective Checklist, MAACL).

El modelo completo original de Brecht y sus colaboradores no se incluyó en el artículo. Se trata de una situación común en los artículos publicados cuando el modelo es extenso y el espacio limitado. Por lo común se reporta el modelo final. Durante el proceso de comprobación del modelo se eliminaron rutas estadísticamente no significativas por medio de la prueba Wald, que es una de dos pruebas más populares para encontrar cuáles parámetros son innecesarios en el modelo (para mayores detalles, véase Bentler, 1995; Ullman, 1996). El modelo final de Brecht y colaboradores se presenta en la figura 35.6 y demuestra algunas de las cuestiones que puede realizar el análisis estructural de covarianza. Por un lado, la estructura del modelo puede probarse a través del tiempo. En otras pala-

FIGURA 35.6 *Modelo estructural de covarianza (datos de Brecht* et al.*)*



bras, es capaz de analizar datos complejos reunidos de manera longitudinal. Los datos o modelo del *momento 1* pueden considerarse como una medida de línea base de las relaciones. Los autores también buscaron la correlación entre el error medido a través del tiempo.

El modelo estructural final fue apoyado por los datos. La χ^2, el estadístico de bondad de ajuste, fue 109.41, con 90 grados de libertad. Dicho valor de chi cuadrada no fue significativo. El índice Bentler-Bonett (IAN) fue .95. Ambos estadísticos indican un buen ajuste de los datos empíricos con el modelo hipotetizado. Todos los coeficientes mostrados en la figura 35.6 fueron estimados por el programa EQS y resultaron estadísticamente significativos.

Los hallazgos de Brecht *et al.* implican que el desarrollo de una mejor relación marital promovería un ajuste psicosocial más sano a la enfermedad, si se acepta el concepto de que se está tratando con un modelado "causal". Con tal información los enfermeros o consejeros cardiacos pueden centrarse en el apoyo de una relación marital sana entre los pacientes cardiacos y sus esposas. Ellos son capaces de enseñar a la pareja estrategias para lograr una relación marital positiva. Al hacerlo así se ofrece al paciente la oportunidad de una mejoría significativa en el estrés emocional.

Conclusiones y reservas

Sería erróneo crear la impresión en la mente del lector de que todos los problemas atacados con el análisis estructural de covarianza o modelamiento de la ecuación estructural funcionan tan bien como los descritos en este capítulo, o que debe utilizarse con todos los problemas de investigación multivariada. Todo lo contrario. El propósito de esta sección final del capítulo consiste en intentar colocar el tema en una perspectiva razonable.

Primero se planteará la pregunta más difícil: ¿cuándo debe utilizarse este procedimiento? Como sucede siempre con dichas preguntas, es difícil decir con claridad y sin ambigüedades cuándo debe utilizarse. Un precepto bastante seguro es que *no* debe utilizarse rutinariamente o para análisis o cálculos estadísticos ordinarios. Por ejemplo, no debe emplearse para analizar factorialmente un conjunto de datos para "descubrir" los factores que están detrás de las variables del conjunto. Sencillamente no se ajusta muy bien al análisis factorial exploratorio y quizá sea un "destructor" al comprobar diferencias de medias entre grupos o subgrupos de datos. Si es posible utilizar un procedimiento más simple —como la regresión múltiple, la regresión logística, las tablas de contingencia multifactoriales o el análisis de varianza— y obtener respuestas a preguntas de investigación, entonces no tiene caso el uso del modelado de la ecuación estructural. Es evidente que se intentará un uso inapropiado. Cada vez es más fácil para los investigadores el uso de los programas computacionales LISREL, EQS y AMOS. Muchos de ellos aseguran en sus anuncios que "no se necesita experiencia para realizar el modelamiento de la ecuación estructural". Lo anterior significa que, entre otras cosas, el modelamiento de la ecuación estructural se utilizará con mayor frecuencia. A diferencia de otros procedimientos, el uso del análisis estructural de covarianza requiere de conceptos bastante difíciles, de la comprensión técnica de la teoría de medición, de la regresión múltiple y del análisis factorial. El rápido acceso a los programas computacionales "fáciles de usar" podría conducir al uso inapropiado del análisis estructural de covarianza. Lo mismo sucedió, aunque en menor grado, con el uso del análisis factorial. Aun así, el análisis factorial ha sido integrado "exitosamente" al cuerpo de la metodología de la investigación del comportamiento, aunque con frecuencia se utiliza inadecuadamente (véase Comrey, 1978). La naturaleza del software mismo hace que esto sea casi inevitable. Uno de sus propósitos es facilitar lo que en esencia no es fácil. Por consiguiente, se verá la publicación de muchos estudios que usarán LISREL, EQS, AMOS y otros programas similares, de forma inadecuada. En síntesis, dichos programas de estructuras de covarianza sólo deben utilizarse en una etapa relativamente tardía de un programa de investigación, cuando se requieran pruebas "cruciales" de hipótesis complejas.

El análisis estructural de covarianza se adecua mejor al estudio y análisis de modelos teóricos estructurales complejos, donde se utilicen cadenas complejas de razonamientos para ligar la teoría con la investigación empírica. Bajo ciertas condiciones y limitaciones, el sistema es un medio poderoso de comprobación de explicaciones alternativas de fenómenos de comportamiento. Resolver un problema de estructuras de covarianza de manera adecuada por lo general requiere de gran cantidad de ideas y análisis preliminares —lejos de la computadora y sus programas—.

Otro uso que bien puede dársele a los programas computacionales de estructuras de covarianza es la verificación de resultados complejos, surgidos a partir de otros análisis. En el pasado, por ejemplo, el análisis de ruta se ha utilizado para analizar los datos de muchos problemas de investigación. A pesar de que el análisis de ruta es un modelo útil para los problemas de investigación —es particularmente útil en la conceptualización de los problemas— no puede lograr lo que hacen programas computacionales como el LISREL. Maruyama y Miller (1979) señalaron esto cuando explicaron por qué utilizaron el LISREL para reanalizar los datos de disgregación de Lewis y St. John (1974). Los programas de modelamiento de la ecuación estructural como el EQS y el LISREL con frecuencia tienen la capacidad de establecer limpiamente aspectos de hipótesis de investigación que otros métodos no logran. Aun así, no es una metodología aplicable de manera general. En definitiva, no se trata de una panacea para estudios mal diseñados.

Con frecuencia existen dificultades técnicas al utilizar tales métodos. Ya se han comentado χ^2 grandes y significativas con grandes cantidades de sujetos, y se han sugerido

remedios, especialmente el estudio de residuales y el uso de otros índices de bondad de ajuste, como el IAN de Bentler o el IAC, los cuales no dependen del tamaño de la muestra. Otro remedio consiste en la comprobación de hipótesis alternativas cuando el problema permite dicha comprobación.

Uno de los problemas más difíciles es el de la *identificación*. Un modelo que se prueba requiere estar sobreidentificado. Lo anterior quiere decir que deben existir más puntos de datos, por lo general varianzas y covarianzas, que los parámetros estimados. Si hay n variables dependientes medidas, entonces no puede haber más de t parámetros estimados a partir de los datos, donde $t = n(n + 1)/2$. Si $n = 5$, entonces $t = 5(5 + 1)/2 = 15$, y no más de 15 parámetros pueden estimarse en un modelo. Existen otras condiciones que pueden hacer que un modelo no sea identificable, pero es extremadamente difícil especificarlas de antemano.

La dificultad técnica más común está íntimamente relacionada con la identificación. Por cualquier razón o combinación de razones, el programa computacional quizá no corra y anuncie que "algo" anda mal. ¿Pero qué es? Por otro lado, la ejecución de la computadora puede haberse completado y alguno de los parámetros tal vez no tenga sentido. Por ejemplo, es posible que se reporten varianzas negativas. ¿Por qué? Cualquiera que haya utilizado programas de computación "enlatados" en cualquier grado está familiarizado con los tétricos mensajes que presenta la computadora. Cuando se consulta a un experto, la respuesta es invariable: "Hay algo que está mal en el modelo." ¡Sí, claro! ¿Pero qué es? Y, naturalmente, los modelos teóricos con frecuencia no se ajustan: "¡Hay algo que está mal en el modelo!" Y también a menudo ocurre el análisis por computadora que funciona magníficamente, pero los estadísticos indican que el modelo del investigador no se ajusta. ¿Está mal la teoría? Si se está fuertemente comprometido con una posición teórica, quizá sea difícil admitirlo. En cualquier caso se deben verificar diversas posibilidades. Primero, el modelo no se ajusta porque fue conceptualizado pobre e incorrectamente. Segundo, no se ajusta debido a que el usuario de los programas computacionales cometió un error (o dos o tres) al utilizar el sistema. Tercero, el análisis computacional no funcionará a causa de que existen defectos en los datos (fuerte colinealidad en una matriz de correlación, por ejemplo); y cuarto, el modelo no se ajusta porque la teoría de donde fue derivado está equivocada o es inaplicable.

La medición inadecuada constituye una limitación de gran parte de la investigación del comportamiento. La dificultad técnica para medir variables psicológicas y sociológicas no ha sido apreciada aún por los investigadores en psicología, sociología y educación. No resulta sencillo diseñar pruebas y escalas para medir constructos psicológicos y sociológicos. Tampoco es fácil realizar la investigación psicométrica para establecer la confiabilidad y validez de las medidas utilizadas. Evidentemente es más difícil aun admitir que las propias medidas son deficientes. Consulte la investigación de Comrey sobre personalidad en el capítulo 34. El desarrollo de la *escala de personalidad de Comrey* tomó 15 años de trabajo. Con demasiada frecuencia las medidas de uso común en la investigación del comportamiento se aceptan y utilizan sin cuestionamiento; y pocas veces se cuestionan los supuestos que subyacen a las variables de estudio. Si, por ejemplo, se mide el *autoritarismo*, se asume que parte de la variable latente *autoritarismo* es el *antiautoritarismo* (cualquier cosa que esto sea). Al inicio de este capítulo y en el capítulo 34 se analizó una investigación que surgió a partir del cuestionamiento del supuesto común de que el *conservadurismo* y el *liberalismo* son opuestos lógicos y empíricos. Por desgracia, se ha realizado una serie de estudios —y se han estropeado— debido a que se realizó la medición de actitudes sociales basadas en este supuesto (véase Kerlinger, 1984). De forma similar, otros estudios han sido estropeados, tal vez arruinados, tanto por supuestos incorrectos como por una medición inadecuada.

No es posible construir una metodología analítica, no importa qué tan bien concebida y poderosa sea, para medidas cuya confiabilidad y validez sean insatisfactorias. La validez hecha a partir de supuestos es una amenaza particularmente severa para las conclusiones científicas, debido a que los procedimientos de medición no se cuestionan ni se prueban: se asume su confiabilidad y validez. Surge un análisis factorial pobre de factorizar lo que en efecto es una opción o construcción poco sustentada de pruebas y escalas. De manera similar, existe un uso pobre del análisis estructural de covarianza cuando algunas o todas las medidas utilizadas poseen una pobre base técnica en la teoría psicométrica e investigación empírica. El punto importante debe enfatizarse fuertemente: procedimientos elegantes aplicados a datos pobres, reunidos sin relación con la teoría y con el análisis lógico, no llegan a producir algo con valor científico.

Otra dificultad que enfrentan los usuarios del análisis estructural de covarianza consiste en que el análisis estructural multivariado moderno es bastante diferente de la mayoría de los primeros análisis estadísticos. La preocupación de la estadística clásica era evaluar si las diferencias entre medias observadas (en el análisis de varianza) o las contribuciones conjuntas y separadas de las variables independientes (en un análisis de regresión múltiple) eran estadísticamente significativas. No obstante, en el modelamiento estructural las implicaciones de una teoría se integran en un modelo que refleja la teoría y sus implicaciones: se incluyen variables latentes, se evalúan sus relaciones y efectos, y se somete a la estructura completa de relaciones a una comprobación simultánea. La(s) prueba(s) se basa(n) en la congruencia del modelo hipotetizado con los datos obtenidos. No es de sorprender que los investigadores experimenten fallas lógicas, técnicas y teóricas. De hecho, resulta sorprendente que los modelos sean comprobados exitosamente, dada la complejidad e incluso delicadeza de la tarea.

Sin embargo, no parece existir una alternativa razonable. La ciencia requiere de la formulación de teorías *y* de su comprobación empírica. La ciencia e investigación del comportamiento se enfrentan con explicaciones psicológicas y sociológicas de complejos fenómenos humanos y sociales. Por lo tanto, requieren tanto de teorías complejas, donde los conjuntos de variables observadas y latentes se relacionen entre sí, *como* de métodos complejos para conceptualizar y analizar los datos que se producen por observaciones y mediciones controladas de los conjuntos de variables. Hasta el momento, el análisis multivariado y el análisis estructural de covarianza parecen ser los caminos más promisorios para lograr las metas de las ciencias del comportamiento. El hecho de que van a plantear muchos problemas metodológicos difíciles, e incluso insolubles, es obvio. El hecho de que producirán avances y beneficios tanto teóricos como prácticos ya se ha demostrado en el presente capítulo.

A pesar de las dificultades y de las reservas mencionadas antes, no cabe duda de que el análisis estructural de covarianza y los programas computacionales que lo implementan son sobresalientes, altamente valiosos y son contribuciones útiles a la investigación científica del comportamiento. Se concluye este capítulo señalando que su uso e influencia tendrá efectos fuertes y saludables en el desarrollo de la teoría psicológica y sociológica y en su comprobación, así como en el avance material de la investigación científica del comportamiento en general.

RESUMEN DEL CAPÍTULO

1. El análisis estructural de covarianza también se denomina modelamiento de ecuación estructural (MEE). Durante algún tiempo se llamó modelamiento causal, aun-

que posteriormente se desechó el término, pues la causalidad no puede inferirse a partir de correlaciones.

2. El análisis estructural de covarianza moderno fue introducido por Bock y Bargmann, y fue desarrollado por Joreskog.
3. El análisis estructural de covarianza se considera la forma más elevada del análisis de datos de las ciencias sociales y del comportamiento. Es una compleja combinación de regresión múltiple y análisis factorial.
4. Los diagramas de ruta con frecuencia se utilizan para desarrollar de forma gráfica el modelo estructural que se va a probar. Existen ciertas reglas a seguir cuando se realizan diagramas de ruta, para facilitar su traducción en análisis de programas computacionales.
5. Existen esencialmente tres tipos de variables en el análisis estructural de covarianza:
   - Variables independientes (medidas o latentes)
   - Variables dependientes (medidas o latentes)
   - Medidas de error
6. Las variables latentes con frecuencia se llaman factores o variables sin medir.
7. El programa computacional utilizado por los autores para realizar el análisis estructural de covarianza es el EQS de Bentler.
8. El EQS se constituye (en opinión del segundo autor de este libro) un método que los principiantes comprenden con mayor facilidad.
9. El LISREL de Joreskog también se usa en muchos estudios reportados en revistas científicas. Se considera que el AMOS es otro programa de empleo sencillo.
10. Los programas computacionales que realizan el análisis estructural de covarianza son capaces de efectuar el análisis factorial confirmatorio.
11. La identificación es uno de los problemas que se encuentran en el modelamiento de estructuras de covarianza. El modelo necesita estar sobreidentificado.
12. Para que el modelo funcione apropiadamente es necesario que haya más puntos de datos que parámetros estimados.
13. El análisis estructural de covarianza se utiliza mejor en etapas tardías de investigación, donde el investigador ya ha reunido suficiente información sobre las relaciones entre las variables.
14. La disponibilidad de programas computacionales como el EQS, LISREL y AMOS, y su creciente facilidad de uso, conllevan la posibilidad de malos estudios de investigación.
15. Sin importar la metodología estadística empleada por el investigador, la validez continúa siendo una meta importante en los estudios de investigación.

SUGERENCIAS DE ESTUDIO

1. El tema del análisis estructural de covarianza presupone el conocimiento del álgebra matricial, del análisis factorial y del análisis de regresión múltiple. Casi todos los artículos de Joreskog resultan difíciles de leer. Sus primeros trabajos están contenidos en el libro de Magidson (1979). A continuación se presenta una lista de referencias que exponen el análisis estructural de covarianza. Algunos son bastante fáciles de leer:

Bentler, P. M. (1980). Causal modeling. *Annual Review of Psychology, 31*, 419-456. [Una de las primeras y claramente escritas exposiciones sobre el análisis estruc-

tural de covarianza. Bentler y otros que realizaban investigación en esta área, en esa época, utilizaban el término causal. Ello fue criticado posteriormente por Freedman (1987) (véase la cita de Freedman más adelante).]

Cliff, N. (1987). Comments on Professor Freedman's paper. *Journal of Educational Statistics, 12*, 158-160. [Una actualización sobre su artículo de 1983, con alguna información proporcionada por el artículo de Freedman.]

Freedman, D. A. (1987). As others see us. A case study of causal modeling methods. *Journal of Educational Statistics, 12*, 101-128. [En este artículo se señaló que no es posible realizar inferencias causales a partir del uso de correlaciones. Esto condujo a que muchos evitaran el uso del término *causal* al tratar con el análisis estructural de covarianza. Freedman también afirma que existe una cantidad de supuestos que son difíciles de verificar y que pueden ser falsos en aplicaciones específicas.]

Hayduk, L. A. (1987). Structural equation modeling with LISREL: Essentials and advances. Baltimore: Johns Hopkins University Press. [Un libro bien escrito con excelentes explicaciones sobre el modelo LISREL para el análisis de ecuaciones estructurales. Sin embargo, no es propio para el principiante debido a que requiere del conocimiento del álgebra matricial.]

Loehlin, J. C. (1998). *Latent variable models: An introduction to factor, path, and structural analysis* (3a. ed.). Mahwah, Nueva Jersey: Lawrence Erlbaum. [Proporciona una buena definición del análisis de ruta y del análisis del rasgo latente. Señala las precauciones que debe tomar el investigador en el análisis estructural de covarianza. Se da un fuerte énfasis a los diagramas de ruta al explicar las ecuaciones estructurales. Los temas cubiertos pueden ser aplicados a cualquiera de los muchos programas computacionales para el análisis estructural de covarianza.]

Ullman, J. B. (1996). Structural equation modeling. En B. G. Tabachnick y L. S. Fidell (eds.), *Using multivariate statistics* (3a. ed.). Nueva York: Harper & Row, 709-811. [Un libro popular; el capítulo 14 está bien escrito y cubre los "derechos y reveses" del modelamiento de la ecuación estructural. Proporciona una buena comparación de los programas computacionales que realizan análisis estructural de covarianza. Con excepción del capítulo 14, el material del libro fue escrito enteramente por Tabachnick y Fidell; Ullman es el único colaborador.]

2. A continuación se presentan seis estudios de investigación que utilizaron de forma benéfica el análisis estructural de covarianza:

Holahan, C. J., Moos, R. H., Holahan, C. K. y Brennen, P. L. (1995). Social support, coping, depressive symptoms in a late-middle-aged sample of patients reporting cardiac illness. *Health Psychology, 14*, 152-163. [Con el uso del LISREL, desarrollaron un modelo predictivo de los síntomas depresivos. El artículo contiene la matriz de correlación de nueve variables observables. Es ideal para estudiantes que desean probar el uso de los programas EQS, LISREL o AMOS.]

Keith, T. Z. (1999). Structural equation modeling in school psychology. En C. R. Reynolds y T. B. Gutkin (eds.), *Handbook of school psychology* (3a. ed.). Nueva York: Willey, 78-107. [Un capítulo sobresaliente escrito por uno de los principales metodólogos de investigación en psicología escolar. Keith ofrece un panorama sobre la forma en que el análisis estructural de covarianza o modelamiento de la ecuación estructural se utiliza para manejar estudios no experimentales complejos en psicología escolar. Es de fácil lectura. Altamente recomendable.]

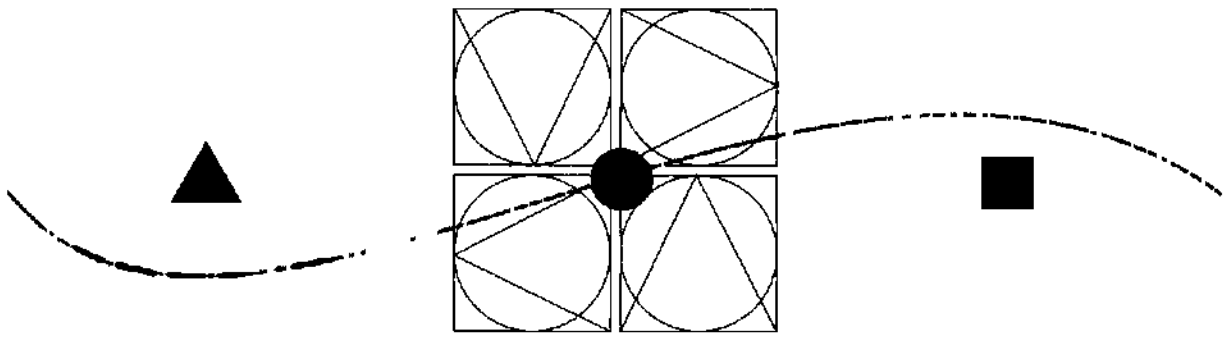
Musil, C. M., Jones, S. L. y Warner, C. D. (1998). Structural equation modeling and its relationship to multiple regression and factor analysis. *Research in Nursing and Health, 21*, 271-281. [Utiliza un modelo conceptual y no técnico para explicar cómo se pueden utilizar las ecuaciones estructurales en la investigación sobre el cuidado del paciente. Los autores muestran cómo el método fue utilizado para estudiar las relaciones entre el estrés, las presiones y la salud física en personas de la tercera edad.]

Nyamathi, A., Stein, J. A. y Brecht, M.L. (1995). Psychosocial predictors of AIDS risk behavior and drug use behavior in homeless and drug addicted women of color. *Health Psychology, 14*, 265-273. [Estos autores desarrollaron un modelo estructural que relaciona recursos personales y sociales, estilos de enfrentamiento, reducción del riesgo y riesgo de SIDA. Obtuvieron un conjunto de factores (variables latentes) y después modelaron dichos factores.]

Wolfle, L. y Robertshaw, D. (1982). Effects of college attendance on locus of control. *Journal of Personality and Social Psychology, 43*, 802-810. [Estudio interesante y bien realizado sobre datos de un estudio longitudinal nacional de la generación 1972 de preparatoria.]

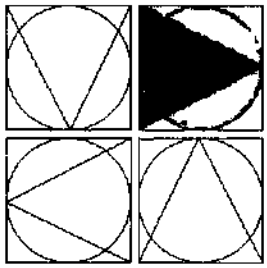
Wyllie, A., Zhang, J. F. y Casswell, S. (1998). Positive responses to televised beer advertisements associated with drinking and problems reported by 18 to 29-year-olds. *Addiction, 93*, 749-760. [Utilizó el modelamiento de la ecuación estructural para estudiar la relación entre respuestas a los anuncios de alcohol y la conducta de beber, y los problemas relacionados con el alcohol. Los investigadores plantearon la hipótesis de que las respuestas positivas a los anuncios de cerveza televisados contribuyen a la cantidad de alcohol consumido en situaciones donde se bebe, lo cual, a su vez, contribuye al nivel de los problemas relacionados con el alcohol. El modelo resultó consistente con la hipótesis.]

Apéndices



Apéndice A

Apéndice B



APÉNDICE A

GUÍA PARA LA ELABORACIÓN DE REPORTES DE INVESTIGACIÓN

El principal medio de comunicación científica es el artículo de investigación. Al paso de los años el formato de dichos reportes se ha estandarizado para cubrir mejor los requerimientos de la comunicación científica. Los convencionalismos de la escritura de un reporte científico se relacionan con la organización del reporte y el estilo de presentación. La elaboración del reporte debe ser tanto breve como clara. Los errores tipográficos, tachaduras y enunciados mal escritos deslucen la presentación del reporte.

Cuando se elabora el reporte de un experimento es necesario que el investigador incluya todo lo que sea relevante al problema de estudio. Deben enfatizarse las bases teóricas del estudio. El lector debe ser capaz de entender la forma en que las predicciones surgen de la teoría. El reporte necesita ser claro en cada detalle en lo que respecta a la manera en que el estudio se realizó. Debe mostrar de forma precisa la manera en que se establecieron las condiciones para permitir la manipulación o el estudio de las variables en el orden demandado por las hipótesis. Debe ser lo suficientemente detallado para permitir que otro investigador independiente replique de manera exacta el estudio. Por último, el reporte requiere establecer qué resultados se obtuvieron y qué interpretación se realizará con ellos, dentro del contexto de la teoría. Un reporte experimental constituye un ciclo completo que inicia en la teoría y termina en la teoría.

Existe una cantidad de estilos populares de elaboración de reportes de investigación. Todos son similares en lo que respecta a lo que el investigador necesita incluir en el artículo. Sin embargo, los detalles dentro de los estilos difieren. Uno de los estilos más populares y "fáciles" es el utilizado por la Asociación Psicológica Americana, el cual con frecuencia se conoce como el "estilo de la APA". A pesar de que originalmente fue desarrollado para las revistas de psicología publicadas por dicha asociación, se ha extendido a revistas que no pertenecen a la APA y también a revistas que no pertenecen al área de la psicología.

Aquí se presentará tal estilo, ya que es el estilo de elaboración que se utiliza con mayor frecuencia en las ciencias sociales y del comportamiento. Inclusive muchas revistas de educación han adoptado este estilo. No obstante, las descripciones serán breves y una de las metas será ofrecer al lector una idea general de la forma en que se organizan los artículos de investigación, lo cual no sustituye al propio manual de la APA (1994). A partir de que surgió como un manual breve, dicho manual ha crecido considerablemente en tamaño y detalle. La edición más reciente, publicada en 1994, consta de 368 páginas. Evidente-

mente, una breve sección en un libro de texto no puede capturar todos los detalles contenidos en el manual completo. La presentación breve incluida aquí será suficiente para brindar al lector un poco de información que ayudará a esa persona cuando consulte el propio manual. Además, existen varias publicaciones dirigidas a ayudar al principiante a aprender y comprender este estilo de presentación (véase Gelfand y Walker, 1990; Hubbach, 1995; Parrott, 1994; Pyrczak y Bruce, 1992). El libro de Hubbach (1995) incluye secciones especiales sobre artículos científicos, el estilo de la APA y ejemplos excelentes.[1]

La estructura del artículo de investigación del estilo utilizado por la APA es:

1. Hoja de presentación con el título
2. Resumen
3. Introducción
4. Método
   a) Participantes
   b) Dispositivos/Materiales
   c) Procedimiento
5. Resultados
6. Discusión
7. Referencias
8. Tablas
9. Figuras

Hoja de presentación

La hoja de presentación es una página separada que contiene el título del estudio. Además contiene el encabezado y el nombre del (de los) autor(es) con su afiliación institucional. El encabezado es una descripción, en una o dos palabras clave, del estudio y aparece en cada página del manuscrito. Si el manuscrito llega a publicarse, el encabezado sirve como una herramienta útil de identificación para el editor.

El propósito del título consiste en proporcionar una descripción en breve del estudio. Para ofrecer la mayor información posible, los títulos por lo general incluyen las variables independientes y dependientes del experimento. Un modelo simple de un título es el siguiente: LAS VARIABLES DEPENDIENTES EN FUNCIÓN DE LAS VARIABLES INDEPENDIENTES. Para anotar un título para un experimento tan sólo se sustituyen la(s) variables(s) dependiente(s) y la(s) variables(s) independientes(s) en el modelo simple, lo cual daría algo como, *La estimulación sexual en función de la cafeína o reducción de la ansiedad por medio del uso de modelamiento videofilmado.* Títulos como "Experimento de psicología" o "Proyecto del curso" NO son aceptables. El título debe contener, por lo general, 15 palabras o menos.

Resumen

El propósito de un resumen es brindar una síntesis del artículo de investigación. Tiene que contener suficiente información para señalarle al lector el propósito y los resultados de la investigación. Debe contener los puntos principales de cada sección del artículo:

- la formulación del problema
- una descripción muy breve del método

[1] El autor agradece a Roberta J. Landi por llamar nuestra atención hacia el libro de Hubbach.

- una definición de todas las abreviaturas y acrónimos
- los resultados más importantes, y
- las conclusiones

El resumen se redacta como un solo párrafo sin separaciones. Al igual que la hoja de presentación, el resumen aparece en una hoja separada. No debe exceder de 15 líneas mecanografiadas a un solo espacio, o de 200 palabras. Tampoco debe incluir ningún dato ni interpretación extensa. Esta sección se intitula "Resumen" y el nombre va centrado en la página.

Introducción

a) La introducción debe iniciar con los *antecedentes* del experimento o estudio. Representa una justificación de la teoría y de las investigaciones previas relevantes al estudio. La introducción le indica al lector la importancia del estudio al ofrecerle una revisión breve sobre la literatura de artículos que son relevantes para el presente estudio de investigación. Si se utiliza el estilo de la APA, ésta es la única sección del artículo que no recibe un título. En otras palabras, en los reportes escritos en el estilo de la APA no aparece un título o encabezado de "Introducción". Se requiere ser preciso al reportar trabajos previos que sean relevantes al estudio de investigación. Cualquier cita textual tiene que escribirse dentro de comillas, anotando la referencia apropiada a la fuente de información. Es importante asegurarse de que los estudios citados sean relevantes al experimento. Las referencias en la introducción (o en cualquier otra sección del reporte) sobre artículos realizados por otros investigadores se realizan anotando el apellido de cada autor y el año de publicación. El año de publicación debe ir entre paréntesis. La referencia completa de cualquier cita se incluye en la sección de "Referencias", que aparece al final del reporte.

Ejemplo

Smith y Martin (1953) reportaron que el desempeño de sus participantes mejoró bajo estas condiciones; mientras que otros observadores notaron un decremento en el desempeño (Burns, 1950; Stevens y White, 1943).

Las referencias específicas respecto a información obtenida de un libro o artículo de revista se indican de la siguiente manera:

Thomas (1983, p. 304) reportó que...
Los resultados de un estudio previo (Carter, 1942, pp. 279-285) condujeron a...

Un investigador nunca debe incluir una referencia que no ha leído personalmente. Para reportar información sobre una fuente secundaria, se cita la fuente secundaria en el texto y se incluye en las referencias.

Ejemplo

En un experimento realizado por Jones y reportado por McGeoch (1952), se encontró que...

b) Una vez que se le han presentado al lector los antecedentes del estudio, la introducción continúa con el propósito o base teórica del experimento. Se establece el

problema específico de estudio junto con una base teórica o literaria de las hipótesis a comprobar y las predicciones y expectativas generales de los resultados de la investigación.

Ejemplo

La disforia puede actuar como una variable importante de confusión... Por lo tanto, se realizó un estudio para examinar las relaciones entre....

c) Después, la hipótesis debe traducirse a términos operacionales. El investigador debe especificar qué variables se manipularán (independientes), en caso de que las haya, y cuáles se observarán (dependientes). Las variables independientes y dependientes deben aclararse *sin* utilizar un enunciado que diga: "La variable independiente fue_____ y la variable dependiente fue_____." En su lugar, se debe decir algo similar a: "En este experimento (estudio), se investigó el número de respuestas correctas a las preguntas de la prueba, en función a la tasa de presentación de las preguntas" o "Se hipotetizó que los efectos indirectos de la relación marital sobre el ajuste psicosocial están mediados por..."

d) Por último, la Introducción se utiliza para definir cualquier término que se utilice por primera vez. Si el investigador desea referirse a las escalas de personalidad de Comrey como EPC, entonces podría escribir una oración como la siguiente:

Las escalas de personalidad de Comrey (Comrey, 1970), que de aquí en adelante serán referidas como EPC, se utilizaron para...

Un término como *aprendizaje* probablemente sea demasiado general para usarse en redacción técnica. Es mejor decir de manera exacta a cuál paradigma de aprendizaje se está refiriendo. Lo mismo se aplica para términos tales como *personalidad* o *ansiedad*.

Existe la tendencia a escribir demasiados detalles en la Introducción. Los detalles del experimento o estudio no deben presentarse en la sección de introducción del reporte. Los detalles se presentan en la sección de Método. Se permite incluir en la Introducción un esbozo o la metodología general seguida; pero sin detalles. Los resultados del estudio no deben aparecer en la Introducción; existe una sección separada para ello.

Método

La sección de Método tiene tres subencabezados: Participantes, Dispositivos y materiales, y Procedimiento. A continuación se describe cada uno de manera separada. La sección de Método, como un todo, describe la experimentación o la conducción del estudio. Tiene que escribirse con suficiente detalle para que otros investigadores puedan, con base en la descripción, repetir de manera *exacta* lo que se hizo.

Participantes

La sección de Participantes consiste en una descripción de las características de los participantes utilizados en el estudio o experimento. Indica quiénes fueron los participantes, cuántos fueron y cualquier detalle que pueda ser relevante. También incluye la manera en que se seleccionaron dichos participantes. Entre las distintas descripciones, la mayoría de

los estudios describe a los participantes en términos del género, edad, nivel educativo, etnia y cualquier otro factor relevante como éstos.

Ejemplo

> Los participantes fueron 48 personas elegidas de una muestra de gente sentada en una fuente ubicada frente al City Hall entre las 10 y 11 AM, un lunes del mes de julio de 1998. Los 18 hombres y 30 mujeres, con edades entre 15 y 35 años, representan cada tercera persona que se detenía en la fuente por un periodo de por lo menos 5 minutos. Otras 10 personas que fueron contactadas se negaron a responder el cuestionario.

Cuando se utilizan grupos de participantes, se debe proporcionar una descripción que indique al lector la forma en que los participantes fueron asignados a los diferentes grupos o condiciones de tratamiento.

Dispositivos y materiales (instrumentación)

Todos los materiales y dispositivos no triviales utilizados en el experimento deben ser descritos en suficiente detalle, para que alguien más pueda establecer una situación idéntica. Si el experimento requirió de lápices, no es necesario definirlos a menos que se trate de lápices raros que tengan un efecto específico en el estudio. Si se utilizan materiales o dispositivos estandarizados, como una prueba de personalidad, por lo común no se incluyen aquí, a menos que tengan algunas características especiales que hayan sido de suma importancia para el experimento. Instrumentos estandarizados ya existentes tales como las escalas de personalidad de Comrey se incluyen en la sección de procedimiento. Si se construyen o utilizan materiales o aparatos nuevos, como aquellos inventados por el investigador, deben describirse de manera completa. Hay que incluir información acerca de los materiales utilizados para medir y/o registrar respuestas. Si se utilizó cierto equipo especial en el experimento, como la "bobina oscilatoria modelo 9 Smith-Johnson" (Smith-Johnson Oscillator Coil Model 9), se requiere informar al lector dónde puede conseguir un recurso como ése.

Procedimiento

La sección de Procedimiento constituye una descripción o explicación de la secuencia de eventos que tuvieron lugar durante la realización del estudio o experimento. En síntesis, indica lo que el (los) experimentador(es) hicieron y lo que sucedió al (los) participante(s). Debe describirse lo que se hizo, en qué orden, durante cuánto tiempo, etcétera.

Ejemplo

> Se obtuvieron los datos de cada participante utilizando un cuestionario sobre la frecuencia del consumo de drogas, durante el cuarto periodo de un día escolar. Se les pidió a los participantes que indicaran la cantidad de alcohol que consumen.

Los métodos estadísticos utilizados para analizar los datos recolectados en el estudio y/o el diseño del estudio de investigación pueden presentarse en la sección de Procedimiento.

Es probable que el investigador descubra que se ha desviado del procedimiento que debía haber seguido; si así sucede, debe describir el procedimiento exactamente como se realizó —y no como tenía que haberse realizado—. Por lo general esto representa un descuido, pero en última instancia llega a disculparse.

Lo que no puede disculparse es la deshonestidad. Las secciones de Procedimiento tienden a complicarse demasiado en experimentos donde existen diversas fases o condiciones. En tal caso, con frecuencia resulta útil adoptar etiquetas para las fases o condiciones. Por ejemplo, con una máquina de enseñanza se podría dividir un ensayo en la "fase de estudio" y la "fase de prueba", con el propósito de hacer referencias posteriores distintivas y simplificadas.

Resultados

En la sección de Resultados se reportan los datos obtenidos en el estudio o experimento, así como los análisis realizados con éstos.

a) Comienza con una descripción de las medidas de la variable dependiente, registradas durante la sesión experimental. Con un ejemplo de la máquina de enseñanza, se registraría el número de errores y el lapso de tiempo ocupado en contestar las preguntas.

b) A continuación, se describen los datos del experimento. Los datos reportados por lo general serán algún tipo de resumen de los datos en bruto. Por ejemplo, tal vez se quiera reportar los resultados en términos de las medias y las desviaciones estándar o los datos en bruto registrados durante la sesión. Se afirmaría: "Se calcularon la media y la desviación estándar del número de errores para cada serie de 20 preguntas."

c) Después se hace referencia a los lugares donde es factible encontrar tales datos. Éstos pueden presentarse ya sea en tablas o en figuras (gráficas). Se les ordena por medio de números y se hace referencia a ellas por su número. Como ejemplo, se podría decir: "La media de los errores de cada serie de 20 preguntas se muestra en la tabla 1 (o figura 1)." Si se tienen diversas variables dependientes, los resultados de cada variable tal vez requieran de una tabla o figura separada. Las reglas para la preparación de tablas y figuras se presentan en una sección posterior de este apéndice.

d) Cuando se hace referencia a una tabla o figura, hay que describir las características importantes de los datos que aparecen en la tabla o figura. En una tabla o figura se presenta mucha información, por lo que es labor del investigador ayudar al lector a comprenderla. Se deben señalar los aspectos importantes, las tendencias generales y cualquier inversión o particularidad que se considere importante; es decir, aquello que parezca ser más que eventos aleatorios.

Es necesario apoyar el análisis de la información con una tabla o figura ofreciendo algunos valores de datos apropiados para ilustrar lo que se desea indicar. Sin embargo, no debe intentarse incluir todos los datos; para eso sirven las tablas y las figuras. Si se tienen 25 páginas de resultados de computadora, dichos resultados deben resumirse y colocarse en una tabla o figura. Si se incluye la hoja de resultados, debe ir como un anexo al documento. Los anexos muy grandes casi siempre son inaceptables si el manuscrito se va a publicar.

Ejemplo

Como se muestra en la figura 1, las funciones para las condiciones con recompensa y sin recompensa inician en el mismo nivel. El número de errores promedio en el primer ensayo fue de aproximadamente 4.5, para ambas condiciones. Después del primer

ensayo, los errores en la condición con recompensa comenzaron a disminuir en una proporción bastante estable, mientras que los errores en la condición sin recompensa permanecieron relativamente constantes. Por ejemplo, en el segundo ensayo se cometieron .50 menos errores en la condición con recompensa, pero para el sexto ensayo dicha diferencia se había incrementado hasta que se cometieron 3.39 menos errores en la condición con recompensa. De forma general, la media del número de errores disminuyó en función de los ensayos en la condición con recompensa; pero no así en la condición sin recompensa.

Una última regla para la elaboración de la sección de Resultados es que *no debe haber discusión* en esta sección. Esto es, no se da una opinión personal o interpretación de los resúmenes de datos. Sólo se presentan los hechos de los hallazgos en la sección de Resultados del reporte. Esto se explica con mayor detalle en la siguiente sección.

Discusión

El propósito de la sección de Discusión consiste en interpretar los resultados y explicar las conclusiones a las que conducen. Es aquí donde se aclara la contribución o valor del experimento o estudio.

a) Por lo común la discusión inicia con una declaración concisa sobre la importancia de los resultados.

Ejemplo

Los resultados del presente estudio coinciden con los de otros estudios que comparan el abuso de alcohol y drogas en jóvenes latinoamericanos.

b) A continuación sigue la interpretación de los resultados. Se hace una inferencia, a partir de las medidas dependientes particulares del experimento, a los procesos psicológicos de interés.

Ejemplo

Los jóvenes asiáticos han presentado de manera consistente menor consumo de drogas debido a que se sienten más amenazados por las consecuencias percibidas por el uso reconocido de drogas.

En apariencia resulta difícil distinguir entre la sección de Resultados y la de Discusión incluidas aquí. En la primera el investigador se adhiere estrictamente a las variables dependientes particulares del estudio. Todas las inferencias, interpretaciones, extrapolaciones y opiniones razonables pertenecen a la sección de Discusión.

Por ejemplo, en la sección de Resultados el estudio de investigación se refiere a la disminución de errores como una función de los ensayos en la condición con recompensa, mientras que los errores permanecen constantes en la condición sin recompensa. En la sección de Discusión, esto podría interpretarse como que la adquisición (o aprendizaje) se llevó a cabo en la condición con recompensa, pero no en la condición sin recompensa. La interpretación de que el aprendizaje se ha visto afectado de manera diferencial involucra una inferencia hecha a partir de los datos de error y, por lo tanto, pertenece a la Discusión y no a la sección de Resultados.

c) Los resultados del experimento o estudio deben, entonces, relacionarse con los resultados de otros estudios sobre problemas iguales o similares, y/o a cualquier teoría relevante con la que se esté familiarizado y que pueda documentarse. Es necesario señalar a qué grado los resultados coinciden o contradicen trabajos previos, qué tanto amplían el cuerpo de conocimientos, qué tanto apoyan o contradicen la teoría, etcétera. La relación de los resultados con otros resultados o teorías también requiere ser analizada. Si existe acuerdo, es suficiente con establecer de manera exacta en qué coinciden. En caso de que no coincidan, se debe ofrecer alguna posible razón de la discrepancia. Por lo general, la primera explicación que se le ocurre al autor es que hubo algún error en el experimento, lo cual puede o no ser verdad. Si es verdad, se debe señalar de forma exacta cuál fue la falla y *por qué*.

d) Cualquier falla o defecto en el experimento que limite la utilidad o generalización de las conclusiones obtenidas necesita discutirse. Cuando se reporta una falla también se debe explicar por qué se trata de una falla e indicar cómo puede corregirse. No es recomendable crear una larga lista de críticas, pues esto sólo creará una mala impresión en el lector o revisor.

e) Una manera adecuada para terminar la discusión consiste en sugerir cuál podría ser el siguiente experimento sobre dicho tema. Si se intenta hacer esto, es necesario asegurarse de explicar el experimento con suficiente detalle para que sea significativo, y se requiere explicar la razón que lo hace el siguiente paso lógico. Frases como las de los siguientes ejemplos deben evitarse ya que consumen tiempo y espacio.

Ejemplos

En el siguiente experimento es necesario utilizar recompensas más grandes.
Se debieron utilizar mejores participantes.

Aquí sería necesario explicar por qué el conjunto actual de participantes fue deficiente, y de qué manera contar con nuevos participantes sería diferente.

Referencias

Únicamente las referencias bibliográficas citadas en el cuerpo del documento deben incluirse en la lista de referencias. *Todas las citas deben aparecer en la lista de referencias.* La lista se realiza en orden alfabético respecto al apellido del primer autor. Después del apellido se anotan la primera y segunda iniciales del autor. *Los nombres de las revistas se escriben completos.*

En el caso de artículos de revistas, se citan los números de las páginas de todo el artículo. En el caso de los libros no se cita el número total de páginas. En un libro publicado, tan sólo se citan las páginas que pertenecen a la parte del libro (capítulo) escrito por el o los autores que se citan. A continuación se presentan ejemplos del estilo utilizado para los tipos de referencias más frecuentes. El tipo de publicación se anota en corchetes sólo para ayudar a identificar cada una; esto no se hace en la sección de referencias del documento.

Note que el estilo de escritura de las referencias de la Asociación Psicológica Americana lleva el año de publicación dentro de paréntesis, después de los nombres de los autores. El título del artículo sigue a la fecha. Sólo la primera letra de la primera palabra de cada oración del título del artículo va con mayúscula, y la primera letra de la primera

palabra después de una coma también se escribe con mayúscula. A continuación se enlista el nombre de la revista, el número del volumen de la revista y los números de las páginas del artículo. El nombre de la revista y el número de volumen se subrayan o se escriben en *itálicas*. Las letras iniciales del nombre de la revista se escriben con mayúsculas (mayúsculas iniciales).

En el caso de los libros, el nombre del autor va seguido por la fecha de publicación en paréntesis. Después se incluye el título del libro. Observe que sólo la primera letra de la primera palabra de cada frase en el título del libro se escribe con mayúscula. El título del libro también se subraya o escribe en *itálicas*. Después del título del libro se incluye el número de edición (2a. ed. [no en itálicas]), o el número de volumen (vol. 3 [no en itálicas]). Si no se requiere anotar una edición o volumen en especial, entonces después del título se escribe un punto. Para concluir, se anota la ciudad seguida por dos puntos (:), y después el nombre de la editorial seguido por un punto (.).

Ejemplos

Erlich, O. y Lee, H. B. (1978). Use of regression analysis in reporting test results for accountability. *Perceptual and Motor Skills, 47*, 879-882. [Artículo de revista, dos autores]

Hollenbeck, A. (1978). Problems of reliability in observational research. En G. Sackett, Ed., *Observing behavior: Vol. 2. Data collection and analysis methods* (pp. 79-98). Baltimore: University Park Press. [Capítulo en un volumen editado, donde el volumen también tiene un título especial]

Jeffrey, W. E. (1969). Early stimulation and cognitive development. En J. P. Hill, Ed., Minnesota Symposia on Child Development (vol. 3) (pp. 46-61). Minneapolis: University of Minnesota Press. [Capítulo en un volumen editado]

Kerlinger, F. N. (1986). *Foundations of behavioral research* (3a. ed.). Fort Worth, TX: Harcourt Brace. [Libro]

Stevens, S. S. (Ed.). (1951). Handbook of experimental psychology. Nueva York: John Wiley & Sons. [Libro con un editor como autor]

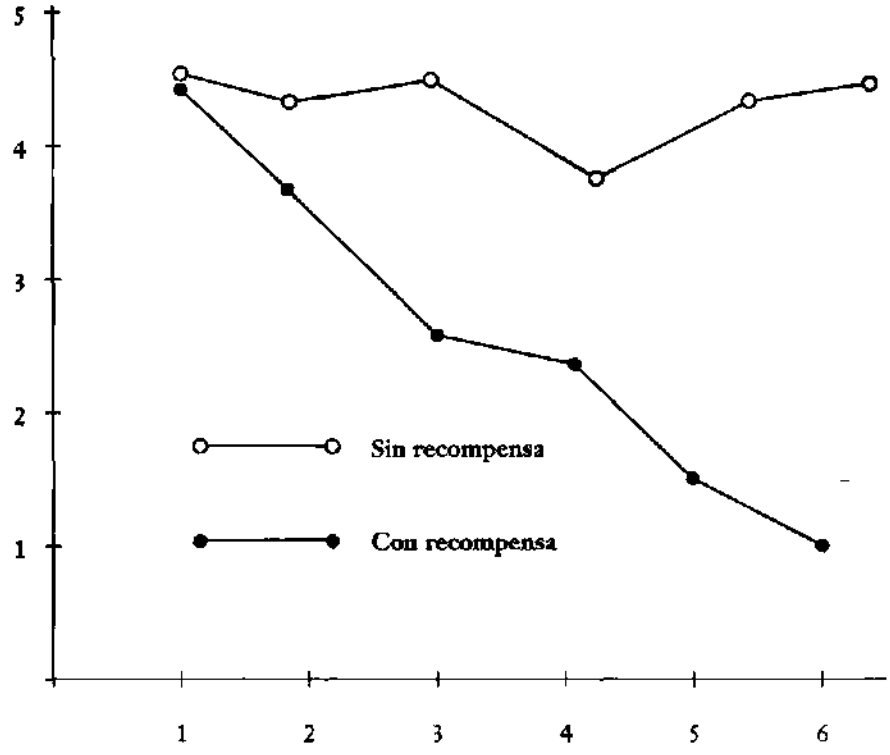
Yi, S. (1977). Some implications of Jeffrey's serial habituation hypotesis: A theoretical basis of resolving one-look versus multiple-look attentional account of discrimination learning. *Journal of General Psychology, 97*, 89-99. [Artículo de revista, un autor]

Preparación de figuras

Las figuras deben contener la información básica necesaria para su comprensión, sin referencia detallada al texto. Lo anterior requiere de la etiquetación cuidadosa de sus partes y un título o pie completo. Cuando se presenta más de una curva en la misma figura, se debe utilizar una leyenda, como en la figura A.1, o poner un nombre directamente a la curva. *El título o pie de la figura aparece debajo de ésta* y consiste de un muy breve resumen de lo que se está graficando. Deben evitarse títulos o pies como "Gráfica de los resultados" o "Una gráfica de..." Sólo se pone mayúscula a la primera palabra del título o pie y se finaliza con un punto. Las figuras se numeran de manera sucesiva con números arábigos. Se debe utilizar una regla para unir los puntos de datos. Hay que evitar dibujar una curva manualmente; de ser posible, debe utilizarse uno de los múltiples programas computacionales para crear gráficas. Por ejemplo, el Excel de Microsoft es capaz de producir gráficas con buena apariencia que son adecuadas para la presentación en artículos. Algunos programas procesadores de textos, como el Word de Microsoft, también son capaces de producir gráficas. En efecto, existen otros programas muy elaborados para la construcción de gráficas.

Ejemplo

FIGURA A.1 *Número promedio de errores en función a los ensayos, bajo condiciones de recompensa y sin recompensa.*



Preparación de tablas

Al igual que con las figuras, las tablas deben contener suficiente información para comprenderse de forma cabal, independientemente del texto. El título debe establecer de forma concisa lo que la tabla contiene. El título debe ser lo más específico posible. Se requiere evitar títulos como: "Tabla de datos", "Tabla de resultados" o "Tabla que muestra..." En general, se deben evitar abreviaturas poco comunes. Si es necesario deben explicarlas en una nota al pie de la tabla.

La tabla se ordena de forma que sea de fácil interpretación para el lector. En caso de ser necesario, debe utilizarse más de una tabla. *El título debe ir centrado encima de la tabla.* Los encabezados, el uso de mayúsculas y otras características importantes de una tabla pueden deducirse a partir del estudio de la tabla A.1 de este documento. Observe que programas de procesadores de textos como el Word de Microsoft poseen la capacidad de generar tablas. Es necesario asegurarse de que los datos incluidos en la tabla estén alineados apropiadamente; es decir, los signos de porcentaje, los decimales y las columnas. Esto facilita al lector el observar y seguir los datos.

Ejemplo

TABLA A.1 *Media de errores en función de los ensayos bajo las condiciones con recompensa y sin recompensa.*

	Media de errores	
Ensayos	*Con recompensa*	*Sin recompensa*
1	4.49	4.51
2	3.80	4.30
3	2.62	4.45
4	2.31	3.87
5	1.46	4.37
6	1.12	4.51

Aunque no existe un límite exacto para el número de figuras y tablas, muchas revistas tienen espacio limitado para los artículos y, por lo tanto, piden que se incluyan únicamente las tablas y figuras más necesarias en el documento real. De ser posible un investigador que posea una gran cantidad de tablas, figuras, hojas de resultados por computadora, etcétera, puede ponerlas a la disposición de los lectores interesados. Existe una organización que acepta dichos materiales y que con una cuota nominal los pone a la disposición de aquellas personas interesadas en ver los datos adicionales. Si el investigador elige utilizar este servicio, debe incluirse una nota de pie o referencia al respecto en el documento real. Dicha nota de pie podría ser como la siguiente.

Ejemplo

[1]Las matrices de correlación en las que se basa este estudio están depositadas en el National Auxiliary Publications Service. Revise el documento NAPS No. ___ con ___ páginas de material suplementario del NAPS, c/o Microfiche Publicatiions, 248 Hempstead Turnpike, West Hempstead, NY 11552. Envíe con anticipación, en dólares estadounidenses únicamente, $ ___ para obtener fotocopias o $ ___ para obtener una microficha. Fuera de Estados Unidos y Canadá añada un franqueo de $ ___ o $ ___ por el franqueo de la microficha.

Uso de abreviaturas

Cuando una palabra o término se utiliza con frecuencia en un reporte, puede abreviarse. Las abreviaturas para el (los) participante(s) y el (los) experimentador(es) son estándares y casi siempre se utilizan:

Ejemplos

sujeto = S sujetos = Ss del sujeto = S's de los sujetos = Ss' experimentador = E experimentadores = Es del experimentador = E's de los experimentadores = Es'

Otras abreviaturas no son estándares y deben definirse la primera vez que se utilicen.

Ejemplo

El aparato era un módulo de prueba visual (MPV). El MPV fue programado para... El instrumento de prueba fueron las escalas de personalidad de Comrey, que de aquí en adelante serán referidas como EPC, que consta de ocho escalas.

Estilo, tiempos, etcétera

La actividad que se describe en el reporte de investigación se llevó a cabo con anterioridad y se debe describir en tiempo pretérito. En raras ocasiones se utilizan pronombres personales, así como los deseos, anhelos, conclusiones, etcétera, del experimentador.

Ejemplos

Pobre:	**El experimentador deseaba encontrar...**
Mejorado:	El propósito del estudio fue...
Pobre:	Nosotros decidimos que el experimento mostraba...
Mejorado:	La conclusión del estudio fue...

Hay que evitar el uso excesivo de expresiones entre paréntesis. Existe la tendencia a referirse a las figuras y a las tablas en paréntesis, lo cual debe evitarse. Por ejemplo, no debe decirse: "La media de errores disminuyó en función a los ensayos (tabla 1)." Sin embargo, en oraciones largas la claridad de la lectura puede mejorarse a través del uso de expresiones entre paréntesis; se requiere tener en mente que se escribe para el lector en general, lo cual significa que los fenómenos deben explicarse. Sin embargo, no se debe adoptar la tarea de enseñar al lector sobre ciertas áreas de comprensión básica. Por ejemplo, es bueno referirse a la teoría de aprendizaje de Skinner sin entrar en los detalles de dicha teoría. No obstante, NO DEBE SUPONERSE QUE EL LECTOR conoce el estudio al que se está haciendo referencia.

Referencias

American Psychological Association (1994). *Publication manual for the American Psychological Association*. Washington, DC: Author.

Gelfand, H. y Walker, C. J. (1990). *Mastering APA style: Student's workbook and training guide*. Washington, DC: American Psychological Association.

Hubbach, S. M. (1995). *Writing research papers across the curriculum* (4a. ed.). Fort Worth, Texas: Harcourt Brace.

Parrott, L. (1994). *How to write psychology papers*. Nueva York: Harper-Collins.

Pyrczak, F. y Bruce, R. R. (1992). *Writing empirical research reports*. Los Ángeles, California: Pyrczak Publishing.

Muestra del reporte de una estudiante

Karen Siegel, quien actualmente es estudiante de doctorado en la Escuela de Psicología Profesional de California, en San Diego, escribió el siguiente reporte cuando era estudiante de licenciatura en la Universidad del Estado de California, Northridge. Éste ilustra de forma clara cómo se escribe un artículo utilizando el estilo de la APA. El reporte se reproduce con el permiso de Karen Siegel.

Forma en que el contacto físico y el género psicológico afectan el comportamiento prosocial

Karen Siegel
Universidad del Estado de California, Northridge

Contacto físico, género psicológico y ayuda

Resumen

El presente estudio predijo que el comportamiento prosocial (de ayuda) entre participantes femeninas y el experimentador se incrementaría con un contacto físico casual e intencional de corta duración. También se investigó el efecto del género psicológico de las participantes (andrógino, femenino o masculino) sobre su conducta de ayuda. 40 participantes voluntarias, que eran estudiantes, completaron un cuestionario (el Inventario del papel femenino de Bem; Bem Sex-Role Inventory), que midió su género psicológico, y después recibieron o no contacto físico, en un diseño entre sujetos de 2 × 3. La variable medida consistió en observar si el sujeto ayudaba a levantar lápices que se caían "accidentalmente". Los resultados mostraron que las participantes que recibieron contacto físico mostraron mayor comportamiento de ayuda que aquellas que no lo recibieron, aunque su género psicológico no tuvo ningún efecto.

Manera en que el contacto físico y el género psicológico afectan el comportamiento prosocial

Existe una gran cantidad de literatura respecto al efecto de diversas formas de contacto físico sobre el comportamiento humano y la salud. Muchos autores consideran que el contacto físico constituye una de las formas más básicas y tempranas de comunicación (Frank, 1957; Montagu, 1971), y que resulta crucial para un desarrollo emocional, social y físico sano (Harlow, 1958). Otra faceta de esta manifestación atávica de comunicación no verbal es su propiedad manipulativa. El contacto físico al servicio del control social (definido por Edinger y Patterson, 1983, como "una respuesta más deliberada y propositiva diseñada para promover un cambio en el comportamiento de la otra persona") ha sido examinado en diversos estudios, empleando variadas modalidades. En un estudio de gran influencia realizado en 1973, Henley (1973) indicó que el contacto físico comunica un mensaje de poder y estatus. Ella reportó una asimetría en los contactos físicos intercambiados entre los sexos (los hombres tocan más a las mujeres en público que a la inversa). Nguyen, Heslin y Nguyen (1975) reportaron que los tipos de contacto físico están asociados a los mismos significados tanto en hombres como en mujeres, aunque los sentimientos difieren y la decodificación precisa de un mensaje táctil depende no sólo de la forma en que se transmite, sino también del lugar donde se aplica.

Un análisis posterior que examinó la generalidad de la asimetría entre los sexos del uso de contacto físico intencional reveló complejidades que impiden una compresión simplista de este aspecto. Ciertos hallazgos de Hall y Veccia (1990) difieren de los de Henley, y sostienen que no está claro qué es lo que los diferentes contactos físicos significan para los sexos, y si la dominancia y el estatus pueden explicar los efectos en el sexo. Major, Schmidlin y Williams (1990) exploraron otras de las muchas variantes del tema de la asimetría. Encontraron que los patrones del género en el contacto físico varían de forma marcada por el ambiente y la edad, lo cual subraya la especificidad situacional de los comportamientos relacionados con el género.

En un contexto menos complejo, se han realizado otros estudios que relacionan el contacto físico (intimidad no verbal) y un incremento en la obediencia a requerimientos hechos (Kleinke, 1976). Willis y Hamm (1980) encontraron que el contacto físico es particularmente importante para lograr la obediencia de personas del mismo sexo. Paulsell y Goldman (1984) examinaron la influencia del contacto físico en diferentes partes del cuerpo (hombro, parte superior e inferior del brazo, y mano) sobre el comportamiento de ayuda. Ellos descubrieron que las mujeres cómplices de los investigadores obtuvieron respuestas de ayuda variables; mientras que los hombres cómplices de los investigadores obtuvieron ayuda con poca variación, independientemente de si los participantes habían o no sido tocados.

Los altos niveles de comportamiento de cuidado y apoyo se dan, discutiblemente, en función de lo que, en la cultura norteamericana, se considera como un comportamiento estereotípico femenino. De acuerdo con Bem (1975), el género psicológico del individuo determina su estilo y rango de comportamiento. Un autoconcepto estrechamente masculino puede inhibir el denominado comportamiento femenino (como ser afectuoso y gentil) y a la inversa; mientras que un autoconcepto andrógino, que incorpora pero no excluye atributos femeninos y masculinos, amplía el rango de comportamientos que pueden elegirse de una situación a otra.

En este experimento se estudió la cualidad manipulativa básica del contacto físico en una interacción limitada. Se probó el efecto de un contacto físico inocuo y casual, dentro del contexto de una interacción verbal, en el comportamiento de ayuda de estudiantes universitarias voluntarias. Además, se evaluó la androginia, feminidad y masculinidad psicológicas de las participantes, por medio del inventario del rol sexual de Bem (1974) (Bem's Sex-Role Inventory).

Contacto físico, género psicológico y ayuda

Se conjeturó que la condición que conduciría a la respuesta de ayuda más consistente sería la del sujeto con contacto físico con un perfil femenino y/o andrógino. De manera inversa, se predijo que aquellos sujetos con un perfil masculino y que no recibieran la situación de contacto físico proporcionarían la menor ayuda de todos.

Método

Se utilizó un diseño entre sujetos de 2 × 3 combinado, donde las situaciones con contacto físico y sin contacto físico funcionaron como una variable independiente; y el género psicológico de las participantes, como la otra variable independiente. El comportamiento de ayuda registrado consistió en observar si los participantes ayudaban a recoger lápices dejados caer por el experimentador, inmediatamente después de implementar la variable del contacto físico.

Participantes
40 mujeres estudiantes de la Universidad del Estado de California, Northridge, con edades que iban aproximadamente de los 18 a los 28 años, sirvieron como voluntarios. En un intento por eliminar cualquier variable extraña y posiblemente provocadora de confusión, respecto a las diferencias culturales en la actitud hacia el contacto físico y al espacio corporal personal, se incluyeron únicamente participantes criados en Estados Unidos.

Procedimiento
Este experimento empleó una condición con contacto físico y una sin contacto físico. El "contacto físico" consistió en una ligera palmada en la parte superior del brazo (el área que obtuvo el nivel más alto de comportamiento de ayuda en el estudio de Paulsell y Goldman [1984]).

Se aplicó el inventario del rol sexual de Bem a cada uno de los sujetos, en forma individual, después de indicarles que completaran el cuestionario. El experimentador les entregó un lápiz proveniente de una caja de 20, el cual posteriormente sirvió como accesorio. La mayor parte de los participantes completaron el inventario, que consistía en 60 adjetivos y frases impresas en una sola hoja, en 10 minutos o menos. También se les instruyó para que preguntaran por la definición de cualquier frase o palabra que no les fuese familiar. Después de que el sujeto completaba el cuestionario, el experimentador —quien permanecía en el cuarto todo el tiempo— caminaba hacia el sujeto sentado, recogía la forma y le agradecía por su participación, durante lo cual se daba o no el contacto físico en la parte superior del brazo. Inmediatamente después de esto, el experimentador dejaba caer lápices de la caja y se agachaba rápidamente para recogerlos. El sujeto se levantaba a ayudar (y se le otorgaba una puntuación de 1) o no lo hacía (y se le otorgaba una puntuación de 0). Antes de empezar a completar el inventario de Bem se le avisó al sujeto que permaneciera sentado al terminar, para hacer que la ayuda fuera una elección más accesible y como un intento de aplicar de manera más efectiva la variable del contacto físico.

Resultados

La proporción de participantes que ayudaron y recibieron contacto físico (.708) fue significativamente mayor que aquellos que no fueron tocados (.438), tal y como lo determinó una prueba z de una cola para proporciones ($z = 1.71$, $p < .05$). También se analizó la significancia del género psicológico de esta manera, comparando las puntuaciones del comportamiento de ayuda de las participantes andróginas con las femeninas, las andróginas con las masculinas y las masculinas con las femeninas, de manera separada; también comparando la variable de contacto físico en cada género psicológico, sin encontrar diferencias significativas entre las proporciones.

Una chi cuadrada, que comparó los efectos de la variable de contacto físico con el género psicológico de cada sujeto en su comportamiento de ayuda, no resultó significativa [$\chi^2 = .97$, $p > .05$].

Discusión

Tal y como se predijo, los resultados de este estudio indicaron que el comportamiento de ayuda de las participantes hacia un experimentador del mismo sexo se incrementó con el contacto físico casual, de corta duración, en la parte superior del brazo.

Se utilizaron participantes mujeres por dos razones. En primer lugar, no existía un cómplice hombre y era altamente probable que casi todos los participantes hombres ayudaran a una experimentadora mujer. (Paulsell y Goldman [1984] reportaron que el 90 por ciento de los participantes hombres tocados en la parte superior del brazo por cómplices mujeres, ayudaron a recoger objetos dejados caer por dichas cómplices.) En segundo lugar, el uso de participantes mujeres controló cualquier asimetría posible en el estatus relacionado al género percibido (Henley, 1973).

A pesar de que se tuvo cuidado en eliminar cualquier otra explicación para este hallazgo significativo, cualquier investigación futura de este tipo sería mejor si se diseñara como un estudio doble ciego, ya que las expectativas del investigador pueden transmitirse de manera no verbal, como con la dirección de la mirada (Kleinke, 1977) o el tono de voz (Goldman y Fordyce). (El experimentador mantuvo, lo más que pudo, una similitud en la comunicación que involucra tales expresiones en particular.)

La reciente popularidad del tema del comportamiento prosocial ha producido un número de combinaciones de manipulaciones para predecir el comportamiento de ayuda, con resultados significativos. Major, Schmidlin y Williams (1990) estudiaron el impacto de la edad de los participantes, aunado a la situación en que ocurría el contacto físico, sobre los patrones de género del contacto físico intencional. Otro estudio realizado por Hewitt y Feltham (1982) combinó el lugar del contacto físico (seis puntos a partir de la mano y hasta la espalda) con el género del experimentador y del sujeto. Nguyen, Heslin y Nguyen (1975) analizaron las interpretaciones de diferentes tipos de contacto físico y zonas corporales en hombres y en mujeres.

Este estudio se realizó a la luz del enfoque en el paradigma moderno de la condición (preferible) de la androginia psicológica. La androginia psicológica es la habilidad individual, como lo define Bem (1974), de no ser dicotómico en el rol sexual, sino tanto masculino como femenino, asertivo y productivo, instrumental y expresivo, para tener acceso a una gama completa de comportamientos que incluyen aspectos "masculinos" y "femeninos". Aunque no se encontró una relación significativa entre el comportamiento prosocial y el género psicológico en dicho estudio en particular, ésta puede ser aun una variable de interés en experimentos futuros específicamente diseñados para medir comportamientos que se vean afectados por el género psicológico del sujeto. Una variante podría ser emplear un experimentador femenino con participantes masculinos, medidos de la misma forma que en el presente estudio, pero utilizando una variable medible distinta que idealmente no involucre comportamientos de cortesía. Otras combinaciones del género del experimentador y del sujeto merecen ser examinadas, así como combinar el efecto del género psicológico con otras variables. En cualquier caso, estudios futuros están garantizados.

Referencias

Bem, S. L. (1974). The measurement of psychological androgyny. *Journal of Consulting and Clinical Psychology, 42*, 155-162.

Bem, S. L. (1975). Sex role adaptability: One consequence of psychological androgyny. *Journal of Personality and Social Psychology, 31*, 634-643.

Edinger, E.. A. y Patterson, M. L. (1983). Nonverbal involvement and social control. *Psychological Bulletin, 93*, 30-56.

Frank, L. K. (1957). Tactile communication. *Genetic Psychology Monographs, 56*, 219-225.

Goldman, M. y Fordyce, J. (1983). Prosocial behavior as affected by eye contact, touch, and voice expression. *Journal of Social Psychology, 121*, 125-129.

Hall, J. A. y Veccia, E. M. (1990). More "touching" observations; new insights on men women and interpersonal touch. *Journal of Personality and Social Psychology 59*, 1155-1162.

Harlow, H. F. (1958). The nature of love. *American Psychologist, 13*, 673-685.

Henley, N. M. (1973). Status and sex: Some touching observations. *Bulletin of the Psychonomic Society, 2*, 91-93.

Hewitt, J. y Feltham, D. (1982). Differential reaction to touch by mean and women. *Perceptual & Motor Skills, 55*, 1291-1294.

Kleinke, C. L. (1977). Compliance to requests made by gazing and touching experimenters in field settings. *Journal of Experimental Social Psychology, 13*, 218-223.

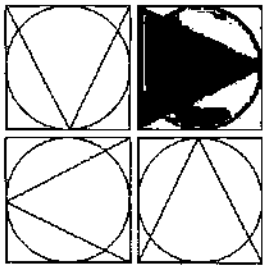
Major, B., Schmidlin, A M. y Williams, L. (1990). Gender patterns in social touch: The impact of setting and age. *Journal of Personality and Social Psychology, 58*, pp. 634-643.

Montagu, A. (1971). *Touching: The human significance of the skin*. Nueva York: Columbia University Press.

Nguyen, T. D., Heslin, R. y Nguyen, M. L. (1975). The meaning of touch: Sex differences. *Journal of Communications, 25*, 92-103.

Paulsell, S. y Goldman, M. (1984). The effect of touching different body areas on prosocial behavior. *The Journal of Social Psychology, 122*, 269-273.

Willis, F. N. y Hamm, H. K. (1980). The use of interpersonal touch in securing compliance. *Journal of Nonverbal Beahavior, 5*, 49-55.



APÉNDICE B[1]

TABLA A *Tabla de números aleatorios*

1	53	95	67	80	79	93	28	69	25	78	13	24	100	62	62	21	11	4	54	44	59	90	78	83	4	97	61	52	75	91
2	62	12	27	41	5	4	19	34	84	78	71	45	73	79	33	57	29	58	75	20	79	78	68	31	25	30	97	31	82	51
3	90	16	47	72	20	60	70	71	2	67	21	65	7	39	58	81	64	11	70	4	79	44	47	7	74	34	55	28	90	19
4	10	59	4	76	80	6	82	20	60	92	33	61	76	83	73	12	84	43	90	71	82	28	21	61	31	92	100	75	22	31
5	32	17	36	64	8	30	80	95	61	33	65	5	39	88	36	44	42	43	5	88	81	13	63	15	47	92	20	62	5	60
6	54	71	27	89	41	53	60	10	2	91	76	95	98	91	64	65	23	57	16	0	90	52	26	90	49	31	68	29	58	10
7	10	60	18	77	34	59	28	99	15	11	70	34	27	78	67	19	97	30	23	60	0	22	11	12	54	50	93	25	69	54
8	42	20	24	36	78	58	82	81	49	91	35	53	30	92	57	19	97	40	58	13	39	42	25	3	97	64	100	55	24	7
9	73	55	87	48	49	97	60	92	27	78	2	55	29	76	99	21	45	72	56	24	16	33	50	84	12	65	4	30	48	56
10	21	56	41	23	58	57	49	49	70	33	6	79	95	3	70	38	26	26	5	89	49	0	68	57	53	91	66	81	53	83
11	9	60	37	99	6	41	69	97	18	44	100	18	46	3	90	57	22	82	15	38	73	97	74	9	35	82	66	34	84	14
12	63	26	41	8	21	38	15	63	38	100	68	89	24	39	19	29	93	97	40	91	70	41	95	83	33	25	33	94	44	39
13	98	72	9	45	69	50	7	86	5	80	0	8	28	96	45	0	0	13	95	24	92	51	11	11	37	91	21	87	89	89
14	87	89	65	22	98	55	86	9	66	43	64	55	80	30	15	99	26	25	71	87	22	39	97	26	50	12	86	22	65	70
15	5	91	68	44	67	2	71	96	15	73	78	3	12	87	53	9	11	12	21	32	57	72	16	35	27	51	91	43	58	61
16	75	93	62	49	95	82	30	81	24	4	11	36	71	96	49	47	65	48	28	8	91	58	40	55	32	7	86	84	95	59
17	76	15	55	38	29	0	8	20	71	42	81	51	44	76	93	42	87	89	38	51	88	65	83	80	66	91	9	68	30	63
18	26	76	93	84	8	40	96	69	84	82	89	5	16	43	34	37	64	39	14	77	95	100	52	99	86	81	65	85	21	9
19	8	35	6	83	76	8	87	81	13	33	14	86	38	23	33	22	58	47	60	36	97	89	20	59	52	9	76	75	52	82

(continúa)

[1] Las tablas estadísticas para la Z, t, χ^2 y F se generaron utilizando los algoritmos computacionales encontrados en las siguientes referencias:

Craig, R. J. (1984). Normal family distribution functions: FORTRAN and BASIC programs. *Journal of Quality Technology, 16*, 232-236.

Kirch, A. (1973). *Introduction to statistics with FORTRAN*. Nueva York: Holt, Rinehart y Winston.

TABLA A *(continuación)*

20	59	73	37	6	26	44	0	24	89	24	78	80	20	8	19	31	32	53	40	32	32	23	57	74	49	17	97	49	71	0
21	87	94	75	45	72	15	39	100	46	99	59	12	22	95	76	18	27	73	88	41	31	99	37	31	24	89	35	14	14	73
22	5	74	8	91	37	5	13	55	13	7	19	24	76	4	25	93	78	9	50	85	98	71	37	53	67	75	9	56	95	71
23	49	82	39	40	51	15	71	53	68	86	50	93	31	22	64	77	46	17	78	25	2	17	69	68	56	44	100	55	80	26
24	2	25	92	97	41	39	98	100	99	67	44	0	99	93	31	69	26	72	56	25	71	42	28	22	96	76	19	63	97	5
25	59	41	49	100	13	0	15	33	82	61	28	59	83	8	17	76	24	58	91	25	3	2	76	87	10	18	23	69	93	27
26	40	13	20	51	81	15	12	45	16	57	47	54	92	60	70	55	98	12	90	27	95	66	23	91	78	86	27	98	16	30
27	80	25	91	36	83	59	19	9	47	61	84	89	98	18	11	56	99	3	26	67	21	24	80	60	44	42	48	77	84	63
28	48	33	7	70	61	95	51	32	89	87	72	6	40	88	52	44	19	96	95	62	12	100	82	5	17	62	65	100	63	9
29	89	5	7	93	48	60	69	97	61	21	87	68	20	4	61	63	75	8	76	92	37	35	40	70	25	86	34	54	53	95
30	97	64	36	36	99	98	23	18	66	28	58	48	34	18	64	71	48	90	63	57	15	14	24	26	65	29	38	85	99	17
31	59	73	71	62	66	34	17	41	32	65	50	73	82	7	20	85	1	65	74	85	23	19	45	61	48	98	84	51	63	70
32	88	75	43	66	66	38	56	31	25	36	26	91	36	100	88	42	74	27	36	40	33	92	18	9	54	51	40	24	82	6
33	34	16	43	38	50	28	34	14	41	2	6	97	56	73	75	17	56	31	100	84	32	25	33	52	26	78	83	44	0	81
34	14	61	81	2	69	73	3	89	79	64	67	80	75	5	66	77	97	30	88	82	52	87	25	63	11	67	93	99	61	39
35	15	39	5	99	29	36	25	40	46	28	34	63	75	18	21	23	13	85	15	43	88	70	92	44	23	73	62	47	60	45
36	68	49	1	55	11	6	63	23	50	33	80	34	82	20	66	48	27	16	86	78	74	89	9	23	66	62	83	28	34	87
37	1	72	18	84	84	86	61	41	22	61	45	36	37	16	20	28	98	36	72	39	67	100	71	8	19	29	0	24	95	26
38	58	73	55	11	9	96	81	84	21	34	50	92	65	91	69	33	23	4	77	93	3	37	95	14	84	27	67	46	61	88
39	91	63	65	63	70	90	57	20	9	13	28	77	72	0	12	30	48	6	28	89	94	6	58	72	73	16	86	19	95	49
40	39	45	31	74	91	85	29	45	98	15	11	50	26	16	36	76	1	40	76	1	88	15	60	27	55	0	83	96	36	53
41	94	12	62	59	14	42	32	75	41	41	0	58	5	78	89	48	35	1	78	70	20	98	38	93	67	35	35	40	38	44
42	3	33	41	22	45	37	65	3	96	27	62	77	16	97	81	78	26	48	94	59	77	82	54	1	63	24	64	31	31	14
43	58	2	83	10	100	50	98	57	32	65	31	87	84	45	0	90	42	78	9	17	21	92	92	47	5	29	6	27	62	72
44	29	73	79	48	66	72	32	1	100	3	2	61	35	0	88	100	45	42	16	18	48	67	36	37	57	12	97	12	95	8
45	55	9	63	66	31	5	8	72	4	85	5	44	4	98	2	79	40	44	96	75	91	59	66	15	41	19	100	33	23	64
46	52	13	44	91	39	85	22	33	4	29	52	6	82	77	25	0	46	100	41	35	46	93	11	9	56	82	97	53	18	86
47	31	52	65	63	88	78	21	35	28	22	91	84	4	30	14	0	97	92	63	87	46	73	55	82	18	76	67	43	76	22
48	44	38	76	99	38	67	60	95	67	68	17	18	46	76	83	5	8	20	87	87	2	42	65	27	16	22	60	18	78	33
49	84	47	44	4	67	22	89	78	44	84	66	15	56	0	90	21	25	88	99	100	32	86	30	50	92	48	55	70	35	20
50	71	50	78	48	65	24	21	24	2	23	65	94	51	82	67	16	35	91	100	35	61	31	75	8	81	58	67	50	28	17
51	42	47	97	81	10	99	40	15	63	77	89	10	32	92	86	32	9	33	79	69	50	7	61	78	15	60	79	47	73	51
52	3	70	75	49	90	92	62	0	47	90	78	63	44	60	13	55	38	64	60	63	92	17	100	2	40	93	83	89	88	20

(continúa)

TABLA A *(continuación)*

53	31	6	46	39	27	93	81	79	100	94	43	39	79	2	18	82	40	30	56	31	81	84	62	41	59	4	46	56	100	58
54	69	27	97	71	52	38	45	35	14	74	40	96	40	88	38	67	44	81	5	12	13	98	21	39	36	74	39	83	77	79
55	2	76	36	72	7	28	55	13	31	78	67	98	50	25	94	39	71	28	0	39	31	69	14	22	50	40	54	12	71	98
56	3	4	20	8	63	33	69	31	69	32	35	18	23	84	69	64	13	43	86	53	10	28	46	41	29	74	46	64	39	4
57	79	55	89	1	25	68	100	58	44	92	73	29	70	47	3	51	37	24	24	29	95	79	80	35	0	9	65	42	99	69
58	99	6	65	35	66	98	66	47	47	22	1	54	94	13	0	31	40	55	69	20	59	12	35	63	52	35	2	56	40	85
59	46	98	1	46	43	86	42	91	63	1	93	84	51	8	79	47	54	85	90	2	19	26	78	95	1	4	72	81	80	60
60	6	14	71	51	7	10	79	41	58	3	27	33	74	67	18	94	4	57	99	37	40	96	68	6	95	55	82	16	36	58
61	92	31	31	40	12	19	74	73	20	94	33	41	40	74	79	42	23	41	29	1	0	13	31	19	63	90	75	17	33	49
62	87	8	68	74	61	66	94	27	71	81	37	82	83	7	8	46	65	63	37	63	88	20	20	75	16	70	26	75	22	48
63	50	48	52	100	68	75	38	65	59	57	78	24	29	52	24	98	78	48	77	64	93	100	50	95	76	94	84	25	67	98
64	67	96	52	88	76	79	16	12	42	33	35	50	54	69	21	57	62	21	84	95	13	66	49	11	48	20	54	51	65	63
65	54	42	22	99	28	90	74	46	26	13	48	45	99	3	38	94	86	53	41	18	35	10	64	79	70	5	55	92	41	92
66	99	51	72	2	75	81	92	71	85	26	77	73	23	14	2	46	7	13	2	40	62	28	72	82	81	51	7	45	9	26
67	35	63	58	46	91	44	56	26	59	56	21	91	19	83	6	61	47	53	10	33	7	97	68	76	44	73	73	0	80	55
68	81	98	63	17	77	45	47	96	25	38	23	26	80	20	47	40	39	14	71	15	60	83	28	56	78	9	27	52	79	68
69	90	47	44	40	40	96	0	62	13	79	39	0	99	57	37	39	2	8	42	58	1	28	1	64	50	28	8	69	70	96
70	29	30	16	54	83	76	50	0	61	100	51	74	78	15	91	61	72	24	44	71	94	59	17	43	50	34	12	14	45	30
71	47	94	70	80	51	26	11	78	34	29	10	55	90	42	4	6	83	72	95	73	24	19	13	98	0	64	44	90	20	13
72	69	14	17	73	79	25	71	14	52	98	77	82	15	25	8	34	38	80	82	97	82	87	98	29	97	69	24	62	100	12
73	54	58	47	9	0	63	6	94	27	3	18	5	36	98	74	36	30	8	87	2	23	76	42	76	87	64	99	5	7	13
74	24	63	57	91	8	58	38	29	72	5	56	71	81	50	67	59	41	9	17	17	85	42	29	80	53	92	6	44	100	18
75	14	24	69	85	97	51	68	80	16	92	59	72	97	23	89	44	16	71	19	83	42	53	54	93	63	19	59	30	80	75
76	86	21	31	59	72	17	77	45	43	29	34	97	67	45	23	88	91	68	12	30	3	41	73	63	76	18	82	8	13	30
77	5	28	80	31	99	77	39	23	69	0	15	49	100	2	22	64	73	92	53	64	7	19	80	64	4	34	30	65	63	11
78	29	71	48	4	87	32	17	90	89	9	99	34	58	8	61	73	98	48	89	90	24	25	98	38	79	45	84	30	49	64
79	90	94	19	80	70	36	2	17	48	63	82	39	85	26	65	27	81	69	83	20	40	25	87	45	88	52	19	33	17	63
80	62	66	48	74	86	6	66	41	15	65	6	41	85	57	84	64	70	39	64	87	62	78	25	71	57	6	98	59	79	34
81	67	54	3	54	23	40	25	95	93	55	59	46	77	55	49	82	26	8	87	54	10	53	29	37	82	5	77	54	4	69
82	75	27	62	15	81	36	22	26	69	42	44	91	55	0	84	48	68	65	5	45	35	11	73	30	16	3	75	56	58	98
83	70	19	7	100	94	53	81	76	73	40	22	58	49	42	96	18	66	89	8	69	17	54	7	86	29	18	86	98	5	56
84	75	7	9	20	58	92	41	42	79	26	91	44	63	87	45	21	23	15	6	72	60	78	88	27	45	80	66	25	37	73
85	55	70	10	23	25	73	91	72	29	47	93	58	21	75	80	52	9	12	36	93	9	58	84	88	90	73	47	49	53	95

(continúa)

Tabla de números aleatorios *(continuación)*

14	94	86	38	11	60	57	16	41	46	20
15	6	62	50	24	11	19	73	14	42	48
16	53	70	54	25	96	38	43	5	2	4
17	28	75	64	90	11	80	94	99	35	54
18	68	57	34	30	29	61	33	49	0	11
19	45	65	89	88	39	93	71	55	29	67
20	73	11	78	58	58	34	20	30	43	40
21	26	59	10	35	75	4	34	38	0	63
22	58	15	70	36	19	49	45	18	36	2
23	87	85	52	76	40	61	50	68	72	7
24	98	44	82	35	0	33	26	68	75	7
25	35	39	8	70	79	48	30	65	65	63
26	79	82	7	23	41	81	8	32	8	8
27	0	30	98	86	100	14	55	86	71	13
28	88	88	48	70	64	81	29	71	62	67
29	45	62	32	83	60	48	0	44	94	22
30	63	8	87	100	28	82	67	65	10	81
31	33	6	49	38	55	78	94	26	4	29
32	79	51	52	9	38	18	13	16	86	42
33	63	29	23	97	64	6	63	74	29	77
34	94	16	38	87	3	25	25	49	22	68
35	32	6	90	100	29	26	31	39	32	93
36	92	99	60	23	79	82	6	62	2	75
37	46	1	2	68	40	8	3	99	19	6
38	65	55	20	58	89	100	74	77	28	30
39	37	58	49	5	51	55	90	22	3	37
40	80	47	63	53	58	95	55	25	67	58
41	2	48	66	86	47	74	48	87	71	21
42	49	71	92	36	55	72	74	13	99	31
43	35	48	56	92	76	75	45	23	91	15
44	77	61	32	6	66	47	66	0	24	26
45	50	83	57	78	38	55	48	97	5	62
46	83	94	8	40	14	39	93	51	42	80

(continúa)

Tabla de números aleatorios *(continuación)*

47	82	1	78	19	94	56	38	8	37	28
48	73	74	13	2	42	64	89	86	72	9
49	54	43	20	13	39	76	59	7	51	19
50	77	32	56	82	56	60	98	80	21	49
51	99	27	39	7	32	7	85	14	22	76
52	1	14	43	75	65	65	63	53	81	57
53	26	51	32	8	24	99	30	36	32	59
54	37	89	4	20	21	91	98	90	37	49
55	25	26	20	61	52	93	90	76	46	19
56	47	55	98	22	69	9	15	34	94	16
57	90	22	16	34	81	44	3	24	96	70
58	2	85	2	58	26	94	48	0	85	70
59	49	67	32	10	28	90	72	25	28	53
60	68	68	69	7	11	31	17	39	82	85
61	13	54	32	26	66	38	1	7	35	16
62	6	1	89	99	21	48	6	9	67	85
63	94	23	75	40	33	86	87	76	24	98
64	33	98	80	13	84	70	85	93	74	22
65	14	63	52	94	56	5	40	55	50	17
66	47	34	47	47	95	45	38	82	85	20
67	84	77	74	27	5	17	57	75	63	2
68	90	48	12	51	55	77	48	10	55	21
69	26	100	6	31	89	0	31	91	5	23
70	79	63	76	72	18	67	87	47	90	93
71	66	81	97	81	11	38	7	37	93	64
72	28	84	86	10	69	25	66	93	21	57
73	33	19	18	37	96	73	95	91	24	24
74	24	31	5	6	37	63	93	42	5	97
75	8	91	48	79	2	40	6	56	57	60
76	78	45	43	77	77	99	98	40	14	82
77	72	20	15	22	30	82	77	51	87	61
78	98	48	25	14	0	12	63	67	12	77
79	60	62	46	12	59	99	5	88	74	89

(continúa)

Tabla de números aleatorios *(continuación)*

80	20	77	87	83	12	74	29	12	16	99
81	7	40	18	32	85	37	73	42	49	49
82	46	93	58	96	29	73	6	71	8	46
83	78	0	78	24	34	73	95	11	44	36
84	7	67	29	27	12	90	60	97	15	94
85	62	28	11	61	0	91	49	32	82	28
86	14	46	52	52	36	21	13	70	24	76
87	26	94	34	57	81	28	49	74	68	50
88	15	11	82	35	77	9	28	11	32	30
89	24	71	92	75	70	60	80	88	21	11
90	18	25	7	100	80	84	97	84	18	53
91	34	91	25	98	77	14	95	100	84	19
92	98	80	72	72	71	66	13	33	24	12
93	22	83	2	33	32	91	78	53	45	63
94	41	39	35	37	66	52	80	1	33	94
95	30	54	73	21	43	68	65	83	26	90
96	65	100	85	12	69	3	72	55	43	5
97	57	69	37	7	62	65	36	9	57	73
98	44	51	38	59	85	91	51	79	14	26
99	39	76	88	46	46	65	72	62	92	67
100	84	60	42	55	48	99	44	66	77	27

	Media	Varianza	Desviación estándar (D. E.)
1	51.8400	895.8144	29.9302
2	46.2000	809.3200	28.4486
3	47.6900	740.6539	27.2150
4	51.8300	872.7611	29.5425
5	53.2100	877.5659	29.6237
6	48.8700	903.9131	30.0651
7	49.6400	778.8704	27.9082
8	51.3700	889.7331	29.8284
9	45.0700	771.7251	27.7799
10	49.2800	872.3016	29.5348
11	48.8700	777.5731	27.8850
12	53.0800	860.2136	29.3294
13	56.5100	773.0099	27.8031
14	47.9900	1110.2299	33.3201

(continúa)

	Media	Varianza	Desviación estándar (D. E.)
15	49.3700	913.6531	30.2267
16	49.0200	714.0396	26.7215
17	45.6800	842.0776	29.0186
18	47.0400	853.3384	29.2120
19	53.5100	977.2499	31.2610
20	52.7400	853.4924	29.2146
21	50.0600	1001.1564	31.6411
22	53.9500	907.7475	30.1288
23	53.6100	737.3779	27.1547
24	49.3100	807.4139	28.4150
25	49.1600	673.9544	25.9606
26	50.2200	855.3316	29.2461
27	58.3600	877.1904	29.6174
28	49.5700	709.7051	26.6403
29	55.4400	868.3664	29.4681
30	49.4300	791.3851	28.1316
31	48.5200	847.9296	29.1192
32	52.9400	802.3564	28.3259
33	46.7900	784.6259	28.0112
34	48.3300	881.4611	29.6894
35	47.2900	759.3059	27.5555
36	35.5100	854.5499	29.2327
37	52.3900	907.8379	30.1303
38	49.9500	851.3275	29.1775
39	46.0000	817.7800	28.5969
40	47.6500	815.1475	28.5508

TABLA B *Tabla de la curva normal (área entre Z = 0 y Z)*

	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0001	0.0040	0.0080	0.0120	0.0160	0.0200	0.0240	0.0280	0.0319	0.0359
0.1	0.0399	0.0438	0.0478	0.0518	0.0557	0.0597	0.0636	0.0675	0.0715	0.0754
0.2	0.0793	0.0832	0.0871	0.0910	0.0949	0.0988	0.1026	0.1065	0.1103	0.1141
0.3	0.1180	0.1218	0.1256	0.1293	0.1331	0.1369	0.1406	0.1444	0.1481	0.1518
0.4	0.1555	0.1591	0.1628	0.1664	0.1701	0.1737	0.1773	0.1809	0.1844	0.1880
0.5	0.1915	0.1950	0.1985	0.2020	0.2054	0.2089	0.2123	0.2157	0.2191	0.2224
0.6	0.2258	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2518	0.2549
0.7	0.7581	0.2612	0.2643	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2882	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3314	0.3339	0.3364	0.3389
1.0	0.3413	0.3437	0.3461	0.3484	0.3508	0.3531	0.3554	0.3576	0.3599	0.3621
1.1	0.3643	0.3664	0.3686	0.3707	0.3728	0.3748	0.3769	0.3789	0.3809	0.3829
1.2	0.3850	0.3869	0.3888	0.3907	0.3926	0.3944	0.3962	0.3980	0.3998	0.4015
1.3	0.4032	0.4049	0.4066	0.4083	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4278	0.4292	0.4305	0.4319

(*continúa*)

TABLA B *(continuación)*

	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
1.5	0.4332	0.4344	0.4357	0.4369	0.4382	0.4394	0.4406	0.4417	0.4429	0.4440
1.6	0.4451	0.4462	0.4473	0.4484	0.4494	0.4504	0.4515	0.4524	0.4534	0.4544
1.7	0.4553	0.4563	0.4572	0.4581	0.4590	0.4598	0.4607	0.4615	0.4623	0.4632
1.8	0.4639	0.4647	0.4655	0.4662	0.4670	0.4677	0.4684	0.4691	0.4698	0.4705
1.9	0.4711	0.4718	0.4724	0.4731	0.4737	0.4743	0.4749	0.4754	0.4760	0.4766
2.0	0.4771	0.4776	0.4782	0.4787	0.4792	0.4797	0.4802	0.4806	0.4811	0.4816
2.1	0.4820	0.4824	0.4829	0.4833	0.4837	0.4841	0.4845	0.4849	0.4852	0.4856
2.2	0.4860	0.4863	0.4867	0.4870	0.4873	0.4877	0.4880	0.4883	0.4886	0.4889
2.3	0.4892	0.4895	0.4897	0.4900	0.4903	0.4905	0.4908	0.4910	0.4913	0.4915
2.4	0.4917	0.4919	0.4922	0.4924	0.4926	0.4928	0.4930	0.4932	0.4934	0.4936
2.5	0.4937	0.4939	0.4941	0.4943	0.4944	0.4946	0.4947	0.4949	0.4950	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4958	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4968	0.4969	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4973	0.4974	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979
2.9	0.4980	0.4981	0.4981	0.4982	0.4982	0.4983	0.4983	0.4984	0.4984	0.4985
3.0	0.4985	0.4986	0.4986	0.4987	0.4987	0.4988	0.4988	0.4988	0.4989	0.4989
3.1	0.4990	0.4990	0.4990	0.4991	0.4991	0.4991	0.4991	0.4992	0.4992	0.4992
3.2	0.4993	0.4993	0.4993	0.4993	0.4994	0.4994	0.4994	0.4994	0.4994	0.4995
3.3	0.4995	0.4995	0.4995	0.4995	0.4996	0.4996	0.4996	0.4996	0.4996	0.4996
3.4	0.4996	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4998

TABLA C *Distribución* t

	Probabilidad					
Grados de libertad	**0.50 0.25**	**0.10 0.05**	**0.05 0.025**	**0.02 0.015**	**0.01 0.00**	de una cola de dos colas
1	1.000	6.34	12.71	31.82	63.66	
2	0.816	2.92	4.30	6.96	9.92	
3	.765	2.35	3.18	4.54	5.84	
4	.741	2.13	2.78	3.75	4.60	
5	.727	2.02	2.57	3.36	4.03	
6	.718	1.94	2.45	3.14	3.71	
7	.711	1.90	2.36	3.00	3.50	
8	.706	1.86	2.31	2.90	3.36	
9	.703	1.83	2.26	2.82	3.25	
10	.700	1.81	2.23	2.76	3.17	
11	.697	1.80	2.20	2.72	3.11	
12	.695	1.78	2.18	2.68	3.06	
13	.694	1.77	2.16	2.65	3.01	
14	.692	1.76	2.14	2.62	2.98	
15	.691	1.75	2.13	2.60	2.95	

(continúa)

TABLA C *(continuación)*

Grados de libertad	0.50 0.25	0.10 0.05	0.05 0.025	0.02 0.015	0.01 0.00	de una cola de dos colas
			Probabilidad			
16	.690	1.75	2.12	2.58	2.92	
17	.689	1.74	2.11	2.57	2.90	
18	.688	1.73	2.10	2.55	2.88	
19	.688	1.73	2.09	2.54	2.86	
20	.687	1.72	2.09	2.53	2.84	
21	.686	1.72	2.08	2.52	2.83	
22	.686	1.72	2.07	2.51	2.82	
23	.685	1.71	2.07	2.50	2.81	
24	.685	1.71	2.06	2.49	2.80	
25	.684	1.71	2.06	2.48	2.79	
30	.683	1.70	2.04	2.46	2.75	
35	.682	1.69	2.03	2.46	2.72	
40	.681	1.68	2.02	2.42	2.71	
60	.678	1.67	2.00	2.39	2.69	
120	.676	1.66	1.98	2.36	2.62	
inf	.674	1.645	1.96	2.33	2.575	

TABLA D *Distribución de la chi cuadrada, cola superior*

gl/α	0.100	0.050	0.025	0.010	0.005	0.001
1	2.71	3.84	5.02	6.63	7.88	10.8
2	4.61	5.99	7.38	9.21	10.6	13.8
3	6.25	7.81	9.35	11.3	12.8	16.3
4	7.78	9.49	11.1	13.3	14.9	18.5
5	9.24	11.1	12.8	15.1	16.7	20.5
6	10.6	12.6	14.4	16.8	18.5	22.5
7	12.0	14.1	16.0	18.5	20.3	24.3
8	13.4	15.5	17.5	20.1	22.0	26.1
9	14.7	16.9	19.0	21.7	23.6	27.9
10	16.0	18.3	20.5	23.2	25.2	29.6
11	17.3	19.7	21.9	24.7	26.8	31.3
12	18.5	21.0	23.3	26.2	28.3	32.9
13	19.8	22.4	24.7	27.7	29.8	34.5
14	21.1	23.7	26.1	29.1	31.3	36.1
15	22.3	25.0	27.5	30.6	32.8	37.7
16	23.5	26.3	28.8	32.0	34.3	39.3
17	24.8	27.6	30.2	33.4	35.7	40.8
18	26.0	28.9	31.5	34.8	37.2	42.3
19	27.2	30.1	32.9	36.2	38.6	43.8
20	28.4	31.4	34.2	37.6	40.0	45.3
21	29.6	32.7	35.5	38.9	41.4	46.8
22	30.8	33.9	36.8	40.3	42.8	48.3

(continúa)

TABLA D *(continuación)*

gl/α	0.100	0.050	0.025	0.010	0.005	0.001
23	32.0	35.2	38.1	41.6	44.2	49.7
24	55.2	36.4	39.4	43.0	45.6	51.2
25	34.4	37.7	40.6	44.3	46.9	52.6
30	40.3	43.8	47.0	50.9	53.7	59.7
35	46.1	49.8	53.2	57.3	60.3	66.6
40	51.8	55.8	59.3	63.7	66.8	73.4
60	74.4	74.4	83.3	88.4	92.0	99.6
80	96.6	101.9	106.6	112.3	116.3	124.8
100	118.5	124.3	129.6	135.8	140.2	149.4

TABLA E *Valores críticos de F (nivel .05 sin negritas, nivel .01 en negritas)**

		Grados de libertad (numerador)									
		1	2	3	4	5	6	7	8	9	10
Grados de libertad (denominador)	1	161.00	200.00	216.00	225.00	230.00	234.00	237.00	239.00	241.00	242.00
		4 052.0	**4 999.0**	**5 403.0**	**5 625.0**	**5 764.0**	**5 859.0**	**5 928.0**	**5 981.0**	**6 022.0**	**6 056.0**
	2	18.51	19.00	19.16	19.25	19.30	19.33	19.36	19.37	19.38	19.39
		98.49	**99.00**	**99.17**	**99.25**	**99.30**	**99.33**	**99.36**	**99.37**	**99.39**	**99.40**
	3	10.13	9.55	9.28	9.12	9.01	8.94	8.88	8.84	8.81	8.78
		34.12	**30.82**	**29.46**	**28.71**	**28.24**	**27.91**	**27.67**	**27.49**	**27.34**	**27.23**
	4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96
		21.20	**18.00**	**16.69**	**15.98**	**15.52**	**15.21**	**14.91**	**14.80**	**14.66**	**14.54**
	5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.78	4.74
		16.26	**13.27**	**12.06**	**11.39**	**10.97**	**10.67**	**10.45**	**10.29**	**10.15**	**10.05**
	6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06
		33.74	**10.92**	**9.78**	**9.15**	**8.75**	**8.47**	**8.26**	**8.10**	**7.98**	**7.87**
	7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.63
		12.25	**9.55**	**8.45**	**7.85**	**7.46**	**7.19**	**7.00**	**6.84**	**6.71**	**6.62**
	8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.34
		11.26	**8.65**	**7.59**	**7.01**	**6.63**	**6.37**	**6.19**	**6.03**	**5.91**	**5.82**
	9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.13
		10.56	**8.02**	**6.99**	**6.42**	**6.06**	**5.80**	**5.62**	**5.47**	**5.35**	**5.26**
	10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.97
		10.04	**7.56**	**6.55**	**5.99**	**5.64**	**5.39**	**5.21**	**5.06**	**4.95**	**4.85**
	11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.86
		9.65	**7.20**	**6.22**	**5.67**	**5.32**	**5.07**	**4.88**	**4.74**	**4.63**	**4.54**
	12	4.75	3.88	3.49	3.26	3.11	3.00	2.92	2.85	2.80	2.76
		9.33	**6.93**	**5.95**	**5.41**	**5.06**	**4.82**	**4.65**	**4.50**	**4.39**	**4.30**
	13	4.67	3.80	3.41	3.18	3.02	2.92	2.84	2.77	2.72	2.67
		9.07	**6.70**	**5.74**	**5.20**	**4.86**	**4.62**	**4.44**	**4.30**	**4.19**	**4.10**
	14	4.60	3.74	3.34	3.11	2.96	2.85	2.77	2.70	2.65	2.60
		8.86	**6.51**	**5.56**	**5.03**	**4.69**	**4.46**	**4.28**	**4.14**	**4.03**	**3.94**
	15	4.54	3.68	3.29	3.06	2.90	2.79	2.70	2.64	2.59	2.55
		8.68	**6.36**	**5.42**	**4.89**	**4.56**	**4.32**	**4.14**	**4.00**	**3.89**	**3.80**

(continúa)

TABLA E *(continuación)*

	Grados de libertad (numerador)									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
16	4.49	3.63	3.24	3.01	2.85	**2.74**	**2.66**	2.59	2.54	2.49
	8.53	**6.23**	**5.29**	**4.77**	**4.44**	**4.20**	**4.03**	**3.89**	**3.78**	**3.69**
17	4.45	3.59	3.20	2.96	2.81	2.70	2.62	2.55	2.50	2.45
	8.40	**6.11**	**5.18**	**4.67**	**4.34**	**4.10**	**3.93**	**3.79**	**3.68**	**3.59**
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41
	8.28	**6.01**	**5.09**	**4.58**	**4.25**	**4.01**	**3.85**	**3.71**	**3.60**	**3.51**
19	4.38	3.52	3.13	2.90	2.74	2.63	**2.55**	**2.48**	2.43	2.38
	8.18	**5.93**	**5.01**	**4.50**	**4.17**	**3.94**	**3.77**	**3.63**	**3.52**	**3.43**
20	4.35	3.49	3.10	2.87	2.72	2.60	2.52	2.45	2.40	2.35
	8.10	**5.85**	**4.94**	**4.43**	**4.10**	**3.87**	**3.71**	**3.56**	**3.45**	**3.37**
22	4.30	3.44	3.05	2.82	2.66	2.55	2.47	2.40	2.35	2.30
	7.94	**5.72**	**4.82**	**4.31**	**3.99**	**3.76**	**3.59**	**3.45**	**3.35**	**3.26**
23	**4.28**	**3.42**	3.03	**2.80**	**2.64**	2.53	2.45	2.38	232.00	**2.28**
	7.88	**5.66**	**4.76**	**4.26**	**3.94**	3.71	**3.54**	**3.41**	**3.30**	**3.21**
25	4.24	3.38	2.99	2.76	2.60	2.49	**2.41**	2.34	2.28	2.24
	7.77	**5.57**	**4.61**	**4.18**	**3.86**	**3.63**	**3.46**	**3.32**	3.21	**3.13**
26	4.22	3.37	2.98	**2.74**	2.59	2.47	2.39	2.32	2.77	2.22
	7.72	**5.53**	**4.64**	**4.14**	**3.82**	**3.59**	**3.42**	**3.29**	**3.27**	**3.09**
28	4.20	3.34	2.95	2.72	2.56	2.44	2.36	2.29	2.94	2.29
	7.64	**5.45**	**4.57**	**4.07**	**3.76**	**3.53**	**3.36**	**3.23**	**3.12**	**3.03**
29	4.18	3.33	2.93	2.70	2.54	2.43	2.35	2.28	2.22	2.18
	7.60	**5.42**	**4.54**	**4.04**	**3.73**	**3.50**	**3.33**	**3.20**	**3.08**	**3.00**
30	4.27	3.32	2.92	2.69	2.53	2.42	2.34	2.27	2.21	2.16
	7.56	**5.39**	**4.51**	**4.02**	**3.70**	**3.47**	**3.30**	**3.17**	**3.06**	**2.98**
34	4.13	3.28	2.88	2.65	2.49	2.38	2.30	2.23	2.27	2.12
	7.44	**5.29**	**4.42**	**3.93**	**3.61**	**3.38**	**3.21**	**3.08**	**2.97**	**2.89**
38	4.20	3.25	2.85	2.62	2.46	2.35	2.26	2.29	2.14	2.09
	7.35	**5.21**	**4.34**	**3.86**	**3.54**	**3.32**	**3.15**	**3.02**	**2.91**	**2.82**
40	4.08	3.23	2.84	2.62	2.45	2.34	2.25	2.28	2.22	2.07
	7.31	**5.18**	**4.31**	**3.83**	**3.51**	**3.29**	**3.12**	**2.99**	**2.88**	**2.80**
46	4.05	3.20	2.81	2.57	2.42	2.30	2.22	2.14	2.09	2.04
	7.21	**5.10**	4.24	**3.76**	**3.44**	**3.22**	**3.05**	**2.92**	**2.82**	**2.73**
50	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	2.07	2.02
	7.17	**5.06**	**4.20**	**3.72**	**3.41**	**3.11**	**3.02**	**2.83**	**2.78**	**2.70**
60	4.00	3.15	**2.76**	**2.52**	2.37	2.25	2.57	2.10	2.04	1.99
	7.08	**4.98**	**4.13**	**3.65**	**3.34**	**3.12**	**2.95**	**2.82**	**2.72**	**2.63**
70	3.98	3.13	2.74	2.50	2.35	2.23	2.14	2.07	2.01	1.97
	7.01	**4.92**	**4.08**	**3.60**	**3.29**	**3.07**	**2.91**	**2.77**	**2.67**	**2.59**
80	3.96	3.11	2.72	2.48	2.33	2.21	2.12	2.05	1.99	1.95
	6.96	**4.88**	**4.04**	**3.56**	**3.25**	**3.04**	**2.87**	**2.74**	**2.64**	**2.55**
100	3.94	3.09	2.70	2.46	2.30	2.19	2.10	2.03	5.97	5.92
	6.90	**4.82**	**3.98**	**3.51**	**3.20**	**2.99**	**2.82**	**2.69**	**2.59**	**2.51**