



Índices, escalas y tipologías

Lo que aprenderá en este capítulo

Aquí concluiremos el estudio de la medición que emprendimos en los capítulos 5 y 6. Aprenderá la lógica y las destrezas necesarias para elaborar mediciones combinadas a partir de varios indicadores de variables.

Introducción

Comparación de índices y escalas

Elaboración de índices

Selección de reactivos

Relaciones bivariadas entre reactivos

Relaciones multivariadas entre reactivos

Puntuación de los índices

Manejo de los datos faltantes

Validación de los índices

Elaboración de índices del estatus de las mujeres

Elaboración de escalas

Escala de distancia social de Bogardo

Escalas de Thurstone

Escalas de Likert

Diferencial semántico

Escalas de Guttman

Tipologías

Puntos principales

Preguntas y ejercicios de repaso

Proyecto de continuidad

Lecturas adicionales

Introducción

Como vimos en los dos capítulos anteriores, muchos conceptos sociales científicos tienen significados complejos y variados. Hacer mediciones que capten tales conceptos puede ser un reto. En particular, quiero recordarle nuestro estudio de la validez de los contenidos, que atañe al hecho de haber captado todas las dimensiones de un concepto.

Para lograr una cobertura amplia, necesitamos hacer muchas observaciones concernientes a determinado concepto. Así, por ejemplo, Bruce Berg (1989:21) aconseja a quienes realizan entrevistas a fondo que preparen las *preguntas esenciales*, las que se "ajustan a obtener la información concreta deseada". Además, el investigador debe disponer *preguntas adicionales*: "preguntas más o menos equivalentes a algunas de las esenciales, pero redactadas de manera ligeramente distinta".

También con los datos cuantitativos se utilizan varios indicadores. Aunque a veces se puede elaborar un reactivo que capte la variable de interés "Sexo: Masculino Femenino" es un ejemplo simple, otras variables son menos sencillas y requieren varios reactivos para medirlas en la forma adecuada.

Los analistas de datos cuantitativos han concebido técnicas específicas para combinar indicadores en una sola medición. En este capítulo estudiamos la elaboración de **índices y escalas** como mediciones combinadas de variables. Escalas e índices pueden

usarse en cualquier forma de investigación social, pero son más comunes en las encuestas y otros métodos cuantitativos. Al final del capítulo dedicamos una sección breve a las **tipologías**, que son importantes tanto para la investigación cualitativa como para la cuantitativa.

En la investigación social cuantitativa se utilizan mediciones combinadas por varias razones. Primero, a pesar del cuidado que se tenga al diseñar estudios que arrojen mediciones válidas y confiables de las variables, el investigador casi nunca puede desarrollar de antemano indicadores únicos de conceptos complicados. Esto es así sobre todo en cuanto a actitudes y tendencias. Por ejemplo, rara vez puede el investigador idear reactivos simples para los cuestionarios que comprendan los grados de prejuicios, religiosidad, tendencias políticas, enajenación, etc. Con toda probabilidad, usted preparará varios reactivos, de modo que cada uno dé alguna indicación de las variables. Sin embargo, cualquiera puede resultar inválido o poco confiable para muchos entrevistados.

Algunas variables se miden fácilmente con indicadores únicos. Para conocer la edad del interrogado podemos preguntarle "¿cuántos años tiene?". Podemos determinar la circulación de un periódico consultando las cifras del propio diario. El diseño del experimento define con claridad el número de veces que se presenta un estímulo a un grupo experimental. Sin embargo, los científicos sociales, mediante una variedad de métodos de investiga-

ción, a menudo quieren estudiar variables que no tienen indicadores únicos claros e inequívocos.

Segunda, quizá sea deseable emplear una medición ordinal más depurada de la variable, disponiendo los casos en varias categorías ordinales de *muy baja* a *muy alta* en, por ejemplo, la variable *enajenación*. Un solo reactivo puede no tener suficientes categorías para obtener el margen de variación deseado, pero una escala o un índice formado por varios reactivos sí pueden tenerlas.

Por último, los índices y las escalas son instrumentos eficaces para el análisis de datos. Si la consideración de un solo reactivo nos da apenas una indicación aproximada de una variable, el considerar varios reactivos nos dará una indicación más comprensiva y precisa. Por ejemplo, un único editorial periodístico puede darnos algunos indicios de la orientación política del diario, mientras que el examen de varios editoriales nos llevará a una evaluación mejor; pero la manipulación simultánea de varios datos puede ser muy complicada. Los índices y sobre todo las escalas son buenos *instrumentos de reducción de datos*: resumen varios indicadores en una sola puntuación numérica al tiempo que, muchas veces, casi conservan los detalles específicos de cada indicador.

Comparación de índices y escalas

En la bibliografía de la investigación social, los términos *índice* y *escala* se usan en forma imprecisa e intercambiable. Antes de considerar las distinciones que haremos en este libro entre índices y escalas, veamos qué tienen en común.

Tanto escalas como índices son mediciones ordinales características de las variables. Ambos ordenan las unidades de análisis según variables concretas, como *religiosidad*, *enajenación*, *nivel socioeconómico*, *prejuicios* o *nivel intelectual*. Por ejemplo, la puntuación de una persona en una escala o índice de religiosidad indica de manera relativa su religiosidad en comparación con otras personas.

En este libro, tanto escalas como índices son *mediciones combinadas de variables*, es decir, mediciones basadas en más de un reactivo. Así, la puntuación de un entrevistado en un índice o es-

cala de religiosidad estaría determinada por las respuestas concretas a varios reactivos del cuestionario, donde cada uno de ellos indicaría su religiosidad. Del mismo modo, la puntuación del CI de una persona se basa en las respuestas a un gran número de preguntas en la prueba. La orientación política de un periódico estaría representada por una puntuación en un índice o escala que refleje las tendencias de sus editoriales sobre varios asuntos políticos.

En este libro distinguiremos índices y escalas según la manera en que se asignan las puntuaciones. Un *índice* se elabora por la simple acumulación de las puntuaciones que se asignen a los atributos. Por ejemplo, para medir los prejuicios sumariamos los enunciados prejuiciosos con los que está de acuerdo cada entrevistado. Por su parte, una *escala* se forma con la asignación de puntuaciones a *esquemas* de respuestas, reconociendo que algunos reactivos reflejan un grado relativamente bajo de la variable mientras que otros reflejan uno mayor, por ejemplo, aceptar que "las mujeres son diferentes que los hombres" es sin duda una prueba pobre de sexismo comparada con "no se debe permitir votar a las mujeres". Así, las escalas aprovechan cualquier *estructura de intensidad* que haya entre los atributos. Un ejemplo más aclarará esta distinción.

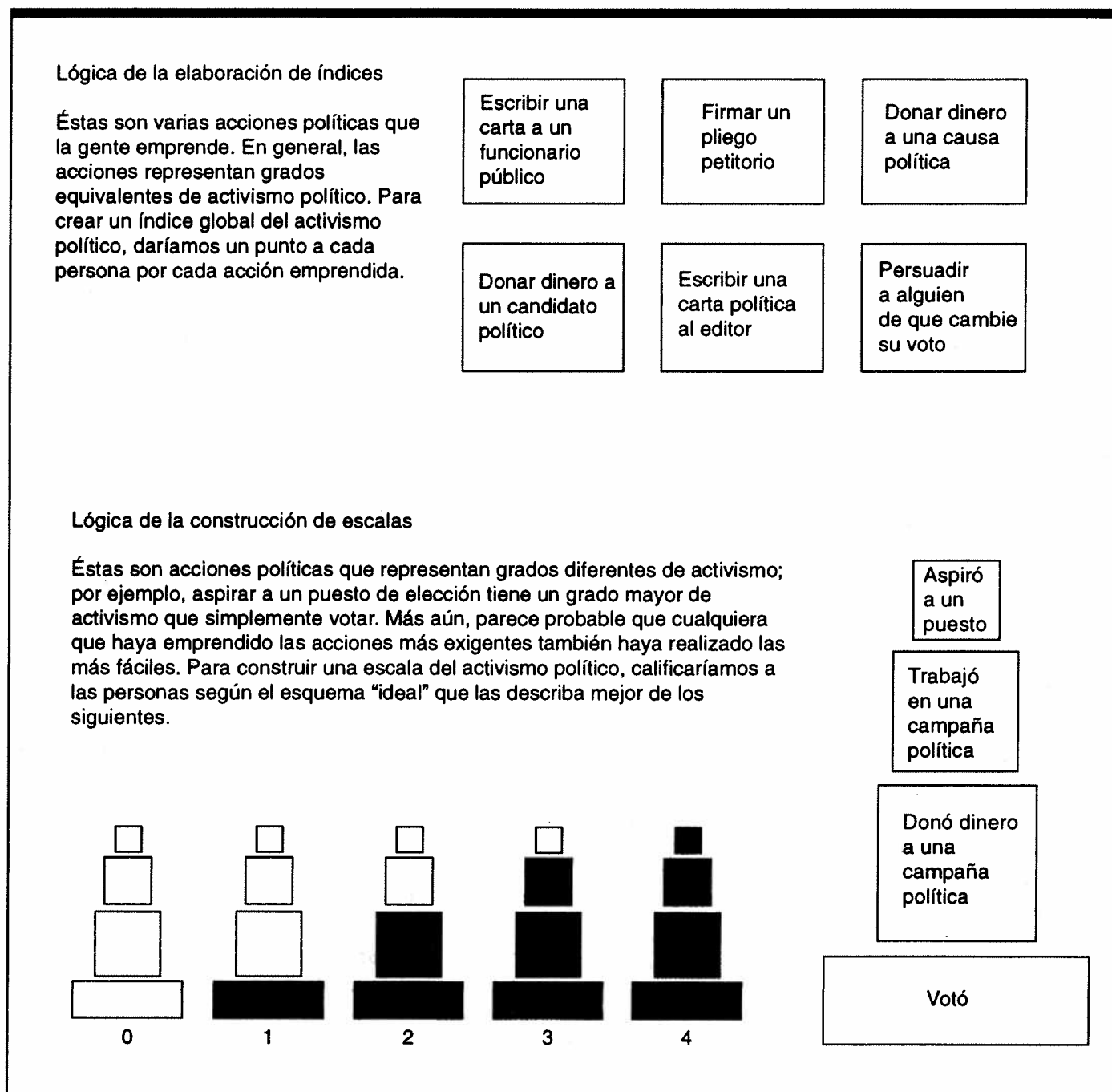
La figura 7.1 ofrece una ilustración gráfica de la diferencia entre índices y escalas. Supongamos que queremos elaborar una medida del activismo político que distinga a los que son muy activos en los asuntos políticos, a los que no participan mucho y a los que se encuentran en alguna postura intermedia.

La primera parte de la figura 7.1 ilustra la lógica de los índices. Están representadas seis acciones políticas. Aunque usted y yo no coincidamos en algunos detalles, creo que estaríamos de acuerdo en que las seis acciones representan aproximadamente el mismo grado de activismo político. Si bien algunas personas preferirían dar dinero que escribir cartas al editor o viceversa, las seis acciones resultan más o menos iguales si consideramos al conjunto de la población.

Podemos construir con las seis acciones un índice de activismo político dando a cada persona un punto por cada acción que emprenda. Así, si usted le escribió a un funcionario público y firmó un plie-

Figura 7.1

Comparación de índices y escalas



go petitorio, tiene en total dos puntos. Si yo aporté dinero para un candidato y persuadí a alguien de que cambiara su voto, tendría la misma puntuación que usted. Con este método, concluiríamos que usted y yo exhibimos el mismo grado de activismo político, aunque hayamos emprendido acciones distintas.

La segunda parte de la figura 7.1 describe la lógica de la construcción de escalas. En este caso,

las acciones representan claramente *grados diferentes* de activismo político, que van desde votar hasta buscar un puesto público. Más aún, parece seguro dar por hecho que hay una *esquema* de acciones en este caso. Por ejemplo, es probable que todos los que hicieron un donativo también hayan votado. Esto indica que la mayoría de las personas se encontrará en sólo uno de los cuatro esquemas "ideales" de acción, representados por los cuadrilá-

teros pequeños en la parte inferior de la figura. Más adelante en este capítulo, en nuestro repaso de las escalas, describiremos las formas de identificar a las personas con el tipo que mejor representan.

En general, las escalas son superiores a los índices porque toman en cuenta la *intensidad* con que los reactivos reflejan la variable que se mide. Asimismo, como muestra el ejemplo de la figura 7.1, las puntuaciones de las escalas transmiten más información que las puntuaciones de los índices. Con todo, manténgase atento al frecuente mal uso del término *escala*; es evidente que no se mejora una medición sólo por llamarla “escala” y no “índice”.

También debe precaverse contra otras dos ideas equivocadas sobre las escalas. Primera, que una combinación de los datos de varios reactivos dé por resultado una escala depende casi siempre de la muestra particular de observaciones que se estudie. Debido a que ciertos reactivos pueden formar una escala de una muestra pero no de otras, no debe usted suponer que cierto conjunto de reactivos es una escala porque la formó con una muestra determinada. Segunda, la aplicación de ciertas *técnicas de construcción de escalas* que estudiaremos no garantiza que la creación de una escala sea mejor que el uso de reactivos que antes formaron una.

Un examen de la considerable bibliografía basada en los datos de las ciencias sociales mostraría que los índices se emplean mucho más que las escalas. Sin embargo, irónicamente, la bibliografía metodológica contiene, si acaso, pocos estudios de la elaboración de índices, mientras que abundan los de la construcción de escalas. Hay dos razones para esta disparidad. Primera, se utilizan más los índices porque suele ser difícil y aun imposible construir escalas con los datos disponibles. Segunda, los métodos de elaboración de índices parecen tan obvios y sencillos que no se analizan mucho.

No obstante, la elaboración de índices no es una empresa simple. La incapacidad general de desarrollar técnicas de elaboración de índices ha dado por resultado la creación de muchos índices deficientes en las ciencias sociales. Con esto presente, dedicaremos más de la mitad del capítulo a la elaboración de índices. Una vez que entienda por completo la lógica de esta actividad, usted estará mejor equipado, o equipada, para ensayar la construcción de escalas. De hecho, un índice realizado con cuidado puede convertirse en una escala.

Elaboración de índices

Veamos ahora las etapas de que consta la elaboración de índices: elegir los reactivos posibles, examinar sus relaciones empíricas, combinar varios reactivos en un índice y validarlo. Como no son del todo evidentes, las expondremos con algún detalle. Deberá terminar esta sección con la capacidad de crear una medición combinada que respalde por completo sus análisis subsecuentes.

Selección de reactivos

Validez aparente La primera etapa es escoger los reactivos para la creación de un índice compuesto destinado a medir alguna variable. El primer criterio para la elección de los reactivos que compondrán el índice es la *validez aparente* (o validez lógica). Por ejemplo, si desea medir el conservadurismo político, sus reactivos deben hacer patente que indican esta variable (o su opuesto, el liberalismo). La afiliación partidista sería uno de los reactivos. Si pidiéramos a los entrevistados que aprobaran o desaprobaban las opiniones de un personaje público conservador muy conocido, lógicamente sus respuestas brindarían una indicación más de conservadurismo. Al elaborar un índice de religiosidad, se deben considerar reactivos como la asistencia a la iglesia, la aceptación de ciertos dogmas religiosos y la frecuencia con que eleva sus oraciones; todos estos reactivos parecen ofrecer alguna indicación de la religiosidad.

Unidimensionalidad En la bibliografía metodológica sobre la conceptualización y la medición se destaca la necesidad de la *unidimensionalidad* en la construcción de escalas e índices: una medición compuesta debe representar sólo una dimensión. Así, los reactivos que reflejen la religiosidad no deben incluirse en una medición del conservadurismo político, aunque puedan relacionarse empíricamente.

General o específico Al mismo tiempo, usted debe estar consciente de los matices de la dimensión general que trata de medir. Así, en el ejemplo de la religiosidad, los indicadores mencionados representan *tipos* de religiosidad, participación en el culto, fe, etc. Si usted quisiera concentrarse en la

participación en el culto religioso, debe escoger los reactivos que indiquen este tipo de religiosidad: asistencia a la iglesia, comunión, confesión, etc. Si quisiera medir la religiosidad en una forma más general, incluiría un conjunto equilibrado de reactivos que representaran sus tipos. En última instancia, la naturaleza de los reactivos incluidos determinará la especificidad o la generalidad con que se mide la variable.

Variación Al elegir los reactivos para un índice, también debe interesarle la *variación*. Por ejemplo, si se pretende que un reactivo indique el conservadurismo político, hay que advertir qué proporción de los encuestados identifica como conservadores. Si determinado reactivo no identificó a nadie como conservador o los identificó a todos —por ejemplo, si nadie indicó su aprobación de un político derechista radical—, no será muy útil para elaborar un índice.

Tiene dos opciones para garantizar la variación. Primera, puede elegir varios reactivos cuyas respuestas dividan a la gente por igual en términos de una variable; por ejemplo, aproximadamente la mitad conservadores y la mitad liberales. Ninguna respuesta única justificaría la caracterización de una persona como muy conservadora, pero la que responda como tal en todos los reactivos así debe quedar caracterizada.

La segunda opción es elegir reactivos que difieran en variación. Un reactivo podría identificar a la mitad de los sujetos como conservadores, mientras que el otro señalaría que pocos de los encuestados lo son. Observe que esta opción es necesaria para construir una escala y también es razonable para elaborar índices.

Relaciones bivariadas entre reactivos

La segunda etapa en la elaboración de índices es examinar las *relaciones bivariadas* entre los reactivos cuya inclusión se contempla. Con esto estamos adelantando un análisis que profundizaremos en el capítulo 15; si así lo desea, repase el Panorama holográfico de ese capítulo. La cuestión básica es notar si las respuestas de los entrevistados a una pre-

gunta —por ejemplo, de un cuestionario— nos dan alguna clave de cómo contestarán otras preguntas.

Supongamos que queremos medir el apoyo de los entrevistados a la participación del país en la Organización de las Naciones Unidas. Un indicador de los grados de apoyo sería esta pregunta: “¿Cree que la ayuda económica a la ONU es excesiva justa insuficiente?”

Otro indicador del apoyo a las Naciones Unidas sería esta pregunta: “¿Debe enviar el país contingentes militares para las acciones de pacificación de la ONU? Muy de acuerdo De acuerdo Muy en desacuerdo En desacuerdo.”

Ambas preguntas, *en apariencia*, reflejan grados distintos de apoyo a las Naciones Unidas. Sin embargo, algunas personas opinarían que el país debe dar más dinero pero no enviar tropas; otras, favorecerían el envío de tropas y el recorte del apoyo económico.

Ahora bien, si los dos reactivos reflejan grados de la misma cosa, debemos anticipar que sus respuestas coincidan en lo general; en concreto, que quienes aprueban la asistencia militar deberían inclinarse más a favorecer el apoyo económico que quienes están en contra del respaldo militar. Por su parte, quienes secundan la ayuda económica tenderían más a consentir la asistencia militar que los que la desaprueban. Si se cumplen estas expectativas, decimos que hay una *relación bivariada* entre los dos reactivos.

Tomemos otro ejemplo. Digamos que queremos determinar el grado al que los encuestados piensan que las mujeres deben tener derecho al aborto. Podríamos preguntar: (1) “¿cree que las mujeres deben tener el derecho de abortar si su embarazo es el resultado de una violación?”, y (2) “¿cree que las mujeres deben tener el derecho de abortar si su embarazo pone en riesgo grave su vida?”

Sin duda, algunos encuestados estarían de acuerdo con el reactivo (1) y en desacuerdo con él (2), en tanto que otros harán lo contrario. Si ambas preguntas tocan alguna opinión general de la gente sobre el tema del aborto, las respuestas a los dos reactivos deben estar relacionadas. Es más probable que quienes apoyan el derecho al aborto en caso de violación lo favorezcan en el caso de que esté amenazada la vida de la mujer que aquellos que lo desaprueban en caso de violación. **Este es otro ejemplo de relación bivariada entre dos reactivos.**

Indicadores de “causa” y “efecto”

por **Kenneth Bollen**

Departamento de Sociología, Universidad de Carolina del Norte, Chapel Hill

Como se explica en el texto, muchas veces es sensato esperar que los indicadores de la misma variable guarden una relación positiva; sin embargo, no siempre es así.

Los indicadores deben estar relacionados si son esencialmente “efectos” de una variable. Por ejemplo, para medir la autoestima podríamos pedirle a una persona que indicara si está de acuerdo con los enunciados (1) “soy una persona buena” y (2) “estoy feliz con lo que soy”. Una persona con mucha autoestima estaría de acuerdo con ambas declaraciones, en tanto que otra con baja autoestima no concordaría con ninguna. Como cada indicador depende de o “refleja” la autoestima, esperamos que tengan una correlación positiva. Más en general, los indicadores que dependen de la misma variable deben estar asociados si son mediciones válidas.

Pero esto no ocurre cuando los indicadores son la “causa” y no el “efecto” de una variable. En esta situación, los indicadores se pueden correlacionar en forma positiva o negativa, o bien no relacionarse en absoluto. Por ejemplo, podríamos tomar el sexo y el grupo étnico como indicadores de la variable *expuesto a discriminación*. No ser blanco o ser mujer

aumenta la probabilidad de sufrir actos de discriminación, de modo que ambos factores son buenos indicadores de la variable; pero no esperamos que el grupo étnico y el sexo estén muy asociados.

Asimismo, podríamos medir el *trato social* con tres indicadores: tiempo compartido con los amigos, tiempo compartido con la familia y tiempo compartido con los compañeros de trabajo. Aunque cada indicador es válido, no tienen que mostrar una relación positiva. Por ejemplo, quizá el tiempo que se pasa con los amigos se relaciona inversamente con el tiempo que se dedica a la familia. Aquí, los tres indicadores “causan” el grado de trato social.

Como último ejemplo, podemos medir *exposición al estrés* de acuerdo con el hecho de que la persona acabe de pasar por un divorcio, de padecer la muerte del cónyuge o la pérdida del empleo. Estos tres acontecimientos indican estrés, pero no tienen que estar relacionados.

En suma, es razonable esperar una asociación entre los indicadores que dependen de o “reflejan” una variable, es decir, que son sus “efectos”. Pero si la variable depende de los indicadores —si los indicadores son la “causa”—, su relación será positiva, negativa o ninguna. Por tanto, debemos decidir si los indicadores son causas o efectos de una variable antes de considerar sus relaciones para evaluar su validez.

Para determinar la fuerza relativa de las relaciones entre pares de reactivos, usted debe examinar todas las relaciones bivariadas posibles entre las variables cuya inclusión en un índice esté considerando. Con este propósito se emplean tablas porcentuales y coeficientes de correlación (véase el capítulo 16). El principal criterio para evaluar estas relaciones es su fuerza; sin embargo, la aplicación de este criterio es más bien delicada. En el recuadro titulado “Indicadores de ‘causa’ y ‘efecto’” se examinan algunas de estas sutilezas.

Tenga cuidado con los reactivos que no se relacionan empíricamente: es poco probable que midan la misma variable. Lo más conveniente es des-

cartar cualquier reactivo que no se relacione con otros.

Al mismo tiempo, una relación muy fuerte entre dos reactivos presenta un problema distinto. Si la relación entre dos reactivos es perfecta, sólo es necesario incluir uno en el índice, puesto que abarca completamente las indicaciones que ofrece el otro (este problema se hará más evidente en la siguiente sección).

Veamos un ejemplo que ilustra la prueba de relaciones bivariadas en la elaboración de índices. Una vez realicé una encuesta con los maestros de una escuela de medicina para averiguar las consecuencias de un “punto de vista científico” en la ca-

lidad de la atención que los médicos brindan a los pacientes. La primera intención era determinar si los médicos de inclinaciones más científicas trataban a sus pacientes en forma más impersonal que los otros.

El cuestionario de la encuesta ofrecía varios indicadores posibles del punto de vista científico de los entrevistados. De ellos, tres reactivos proporcionaron indicaciones especialmente claras sobre la inclinación científica de los médicos:

1. Como catedrático de la escuela de medicina, ¿en qué calidad cree que puede hacer su mayor contribución *docente*: como médico en ejercicio o como investigador médico?
2. A medida que aumentan sus conocimientos médicos, ¿diría que su mayor interés médico radica principalmente en la dirección del manejo integral del paciente o en la comprensión de los mecanismos básicos? [El objetivo de este reactivo era distinguir a quienes les interesaba más la atención general al paciente de quienes se interesaban sobre todo en los procesos biológicos.]
3. En el campo de la investigación terapéutica, ¿en general está más interesado en los artículos que examinan la eficacia de diversos tratamientos o en los que exploran las razones que fundamentan las terapias? [Igualmente, quería distinguir a quienes estaban más interesados en la atención de los enfermos de los que se interesaban más por los procesos biológicos.]

(BABBIE, 1970:27-31)

En cada uno de estos reactivos concluiríamos que los que escogieron la segunda respuesta tienen una inclinación más científica que quienes eligieron la primera. Sin embargo, aunque esta conclusión *comparativa* es razonable, no debemos cometer el error de pensar que los entrevistados que optaron por la segunda respuesta en algún reactivo sean científicos en ningún sentido absoluto: simplemente, son *más científicos* que aquellos que eligieron la primera respuesta. Para ver esto más claramente, examinemos la distribución de las respuestas de cada reactivo. En el primero —la mejor función docente—, sólo alrededor de un tercio de los entrevistados mostró una orientación científica (aproximadamente un tercio dijo que haría su mayor contribución docente como investigador médi-

co). En el segundo reactivo —el mayor interés médico—, alrededor de dos tercios escogieron la respuesta científica: les interesaba más comprender los mecanismos básicos que aprender sobre el manejo integral del paciente. En el tercer reactivo —preferencias de lectura—, aproximadamente 80 por ciento eligió la respuesta científica.

Así, estos tres reactivos no nos indican la cantidad de “científicos” que hay en la muestra, porque ninguno se relaciona con un grupo de criterios de lo que constituya ser científico en algún sentido absoluto. Si tomamos los reactivos con este propósito se nos presentaría el problema de tener tres cálculos muy distintos de la cantidad de científicos que hay en la muestra.

Sin embargo, los reactivos nos ofrecen tres indicadores independientes de las inclinaciones *relativas* de los entrevistados hacia la ciencia. Cada reactivo separa a los interrogados *más* científicos de los que lo son *menos*; pero cada agrupamiento de sujetos más o menos científicos tendrá diferentes miembros que los otros. Los encuestados que parecen científicos en términos de un reactivo no lo parecerán en términos de otro. No obstante, en la medida en que cada reactivo mide la misma dimensión general, deberíamos encontrar alguna correspondencia entre los agrupamientos. Los encuestados que se muestran científicos de acuerdo con un reactivo deberían tender más a aparecer como científicos en su respuesta a otro reactivo que quienes no resultaron científicos en su respuesta al primero. Deberíamos encontrar una asociación o correlación entre las respuestas dadas a dos reactivos.

La figura 7.2 muestra las asociaciones entre las respuestas a los reactivos. Se presentan tres tablas bivariadas que muestran la distribución de las respuestas de cada par de reactivos. Aunque cada reactivo da por resultado un agrupamiento distinto de entrevistados “científicos” y “no científicos”, vemos en la figura que las respuestas a cada reactivo corresponden hasta cierto grado con las respuestas que se dieron a cada uno de los otros.

Un examen de las tres relaciones bivariadas que se presentan en la figura 7.2 respalda la proposición de que los tres reactivos miden la misma variable: las *orientaciones científicas*. Para ver la razón, comencemos por la primera relación bivariada de la tabla. Ahí se muestra que los catedráticos que

Figura 7.2
Relaciones bivariadas entre reactivos de orientación científica

A.		Mejor función docente	
		Médico	Investigador
Mayor interés médico	Manejo integral del paciente	49%	13%
	Mecanismos básicos	51%	87%
		100% (268)	100% (159)
B.		Preferencias de lectura	
		Eficacia	Razones
Mayor interés médico	Manejo integral del paciente	68%	30%
	Mecanismos básicos	32%	70%
		100% (78)	100% (349)
C.		Preferencias de lectura	
		Eficacia	Razones
Mejor función docente	Médico	85%	64%
	Investigador	15%	36%
		100% (78)	100% (349)

respondieron que "investigador" sería su mejor función docente tendieron más a señalar como su mayor interés médico "los mecanismos básicos" que los que señalaron "médico". Los datos muestran que 87 por ciento de los "investigadores" también eligieron la respuesta científica en el segundo reactivo, en comparación con 51 por ciento de los "médicos". (Nota: Es irrelevante el hecho de que los "médicos" se distribuyan de manera uniforme en cuanto a su mayor interés médico. Lo único que importa es que su interés médico es menos científico que el de los "investigadores".) La fuerza

de esta relación se resumiría en una diferencia porcentual de 36 puntos.

La misma conclusión general se extrae de las otras relaciones bivariadas. La fuerza de la relación entre las preferencias de lectura y el mayor interés médico se resume en una diferencia de 38 puntos porcentuales, y la fuerza de la relación entre preferencias de lectura y las dos funciones docentes en una diferencia de 21 puntos.

Al principio elegí los tres reactivos por su validez aparente (los tres parecían indicar las orientaciones científicas de los catedráticos). Al examinar la

Figura 7.3

Relaciones bivariadas entre reactivos de orientación científica

Porcentaje interesado en los mecanismos básicos		Mejor función docente	
		Médico	Investigador
Preferencias de lectura	Eficacia	27% (66)	58% (12)
	Razones	58% (219)	87% (130)

relación bivariada entre pares de reactivos, encontramos apoyo a la suposición de que todos medían básicamente lo mismo; sin embargo, este apoyo no basta para justificar la inclusión de los reactivos en un índice combinado.

Relaciones multivariadas entre reactivos

Antes de combinarlas en un solo índice, necesitamos examinar las *relaciones multivariadas* entre diversas variables. Mientras que las relaciones bivariadas atañen a dos variables a la vez, las multivariadas toman más de dos.

Recuerde que el propósito principal de la elaboración de índices es desarrollar un método de clasificar a los sujetos de acuerdo con alguna variable, como *conservadurismo político*, *religiosidad*, *orientación científica* o cualquiera otra. Un índice de conservadurismo político debe identificar a todos los que sean muy conservadores, moderadamente conservadores, no muy conservadores y nada conservadores (o, en lugar de las dos últimas categorías, moderadamente liberal y muy liberal). Las gradaciones de la variable provienen de la combinación de las respuestas dadas a los reactivos que comprende el índice. Así, el entrevistado que aparece como conservador en todos los reactivos debe considerarse muy conservador en general.

Para que un índice ofrezca gradaciones significativas en este sentido, cada reactivo debe añadir algo a la evaluación del entrevistado. Recuerde que en la sección anterior dijimos que no deben incluir-

se en el mismo índice dos reactivos que guardan una relación perfecta. Si se incluye uno, el otro no añadiría nada a nuestra evaluación de los interrogados. El examen de las relaciones multivariadas entre los reactivos es otra forma de quitar la paja; también determina la capacidad general de cierto conjunto de reactivos para medir la variable que se estudia.

Los propósitos del examen multivariado se aclararán más si regresamos al ejemplo anterior de medir las tendencias científicas de los catedráticos de la escuela de medicina. La figura 7.3 presenta las relaciones trivariadas entre los tres reactivos.

Con una presentación un poco distinta que la figura 7.2, la figura 7.3 clasifica la muestra de los entrevistados en cuatro grupos de acuerdo con (1) su mejor función docente, y (2) sus preferencias de lectura. Las cifras entre paréntesis indican el número de entrevistados de cada grupo (así, 66 de los catedráticos que dijeron que enseñarían mejor como médicos también dijeron que preferían los artículos sobre la eficacia de los tratamientos). En cada grupo se indica el porcentaje de los que dijeron que su mayor interés son los mecanismos básicos (de los 66 catedráticos citados, a 27 por ciento le interesan principalmente los mecanismos básicos).

La organización de los cuatro grupos se basa en una conclusión previa sobre sus orientaciones científicas. Suponemos que el grupo de la esquina superior izquierda es el que tiene menor inclinación científica, dadas su mejor función docente y preferencias de lectura. Pensamos que el grupo de la esquina inferior derecha es el de mayor orientación científica en términos de esos mismos reactivos.

Figura 7.4

Relaciones multivariadas hipotéticas entre reactivos de orientación científica

Porcentaje interesado en los mecanismos básicos		Mejor función docente	
		Médico	Investigador
Preferencias de lectura	Eficacia	51% (66%)	87% (12)
	Razones	51% (219)	87% (130)

Recuerde que también consideramos la expresión de un interés mayor en los mecanismos básicos como indicación de las tendencias científicas. Por tanto, como lo esperaríamos, los de la esquina inferior derecha tienden más a dar esta respuesta (89 por ciento) que los de la esquina superior izquierda (27 por ciento). El interés por los mecanismos básicos de los entrevistados que dieron respuestas mixtas en cuanto a sus funciones docentes y preferencias de lectura tiene un nivel intermedio (58 por ciento en ambos casos).

Esta tabla nos revela muchas cosas. Primera, observamos que la presencia del tercer reactivo no influye mucho en la relación original entre pares de reactivos. Por ejemplo, recuerde que la relación entre función docente y mayor interés médico se resumía en 36 puntos porcentuales de diferencia. En la figura 7.3 vemos que sólo entre los entrevistados que están más interesados en los artículos que abordan la eficacia de los tratamientos, la relación de función docente y mayor interés médico es de 31 puntos porcentuales (58 menos 27 por ciento, primera fila), y lo mismo ocurre con quienes están más interesados en los artículos que exponen las razones de los tratamientos (89 menos 58 por ciento, segunda fila). La relación original entre función docente y mayor interés médico es en esencia la misma que en la figura 7.2, incluso entre los interrogados considerados científicos o no científicos según sus preferencias de lectura.

Llegamos a la misma conclusión con las columnas de la figura 7.3. Recuerde que la relación original entre preferencias de lecturas y mayor interés

médico se resumía como una diferencia de 38 puntos porcentuales. Si en la figura 7.3 tomamos sólo a los “médicos”, vemos que ahora la relación entre los otros dos reactivos es de 31 puntos porcentuales. Se aprecia la misma relación entre los “investigadores” de la segunda columna.

La importancia de estas observaciones se hace más evidente si consideramos lo que ocurrió. En la figura 7.4, unos datos hipotéticos cuentan una historia muy diferente que los datos reales de la figura 7.3. Como vemos, la figura 7.4 muestra que persiste la relación original entre función docente y mayor interés médico, aunque se introduzcan en el cuadro las preferencias de lectura. En cada fila de la tabla, es más probable que los “investigadores” expresen un interés en los mecanismos básicos que los “médicos”. Sin embargo, al recorrer las columnas advertimos que no hay ninguna relación entre preferencias de lectura y mayor interés médico. Si ya sabemos que un entrevistado piensa que su mejor función docente es como médico o bien como investigador, el hecho de conocer sus preferencias de lectura no añade nada a nuestra evaluación de sus inclinaciones científicas. Si la figura 7.4 fuera el resultado de datos reales, concluiríamos que las preferencias de lectura no deben incluirse en el mismo índice combinado con las funciones docentes, pues nada le aportan.

Este ejemplo tiene sólo tres reactivos de cuestionario. Si se contemplaran más, se necesitarían tablas multivariadas más complicadas, compuestas de cuatro, cinco o más variables. Repitamos que el propósito de esta etapa de la elaboración de índices

es descubrir las interacciones simultáneas entre los reactivos para elegir los que han de incluirse en el mismo índice.

Puntuación de los índices

Una vez elegidos los mejores reactivos para el índice, se asignan puntuaciones a las respuestas para crear un índice único combinado de varios reactivos. En esta etapa se deben tomar dos decisiones básicas.

Primera, hay que decidir cuál es el margen deseado de las puntuaciones del índice. Desde luego, una de las principales ventajas de un índice sobre un solo reactivo es el margen de gradaciones que ofrece en la medición de una variable. Como señalamos, el *conservadurismo político* puede medirse de “muy conservador” a “nada conservador” (o “muy liberal”). Así, ¿cuánto deben extenderse los extremos del índice?

En esta decisión surge de nuevo el asunto de la variación. Casi siempre, a medida que los extremos del índice se extienden, en cada uno se encuentran menos casos. El investigador que quiere medir el conservadurismo político hasta su mayor extremo encontrará que no hay casi nadie en esa categoría.

Por tanto, esta primera decisión atañe al deseo conflictivo de tener (1) un margen de medición en el índice y (2) un número adecuado de datos en cada punto del índice. Uno se obliga así a alcanzar alguna clase de equilibrio entre las dos aspiraciones contradictorias.

La segunda decisión concierne a la asignación real de puntuaciones a cada respuesta. Básicamente, se trata de decidir si debe dar el mismo o diferente peso a cada reactivo en el índice. Aunque no hay reglas firmes, le sugiero —y la práctica tiende a sustentar este método— que asigne el mismo peso a los reactivos a menos que haya razones imperiosas para hacerlo de otro modo. Es decir, la carga de la prueba recaería en los pesos diferenciales; los pesos iguales deben ser la norma.

Desde luego, hay que enlazar esta decisión con la cuestión citada del equilibrio entre los reactivos elegidos. Si el índice va a representar la combinación de aspectos ligeramente distintos de la variable, debe darle a cada aspecto el mismo peso. Sin

embargo, en algunos casos se puede pensar que, digamos, dos reactivos reflejan en esencia el mismo aspecto y otro más un aspecto diferente. Si usted quiere representar por igual ambos aspectos en el índice, podría darle al tercer reactivo un peso equivalente al de la combinación de los dos parecidos. En tal situación, conviene asignar un máximo de 2 puntos al reactivo diferente y un máximo de 1 a cada uno de los similares.

Aunque al reflexionar en la puntuación de las respuestas hay que tomar en cuenta estas preocupaciones, por lo general uno experimenta con varios métodos de puntuar, examinando los pesos relativos dados a los aspectos, pero sin dejar de ver al mismo tiempo el margen y la distribución de casos. En última instancia, el método de puntuación que elija representará un equilibrio entre estas diversas exigencias. Desde luego, al igual que en la mayoría de las actividades de investigación, esta decisión está sujeta a revisiones. La validación del índice, que estudiaremos dentro de poco, puede llevarlo a recomenzar sus esfuerzos y elaborar un índice completamente distinto.

En el ejemplo tomado de la encuesta entre los catedráticos de la escuela de medicina, decidí dar el mismo peso a los reactivos, puesto que en parte los escogí porque representaban aspectos ligeramente diferentes de la variable general *orientación científica*. En cada uno de los reactivos, los entrevistados acumulaban un punto por elegir la respuesta “científica” y ninguno por la otra. Por tanto, podían recibir una calificación de 0, 1, 2 y 3. Este método de puntuación me dio lo que consideré un margen útil de variación —cuatro categorías para el índice—, y también suficientes casos para análisis en cada categoría.

Veamos un ejemplo parecido de puntuación en un índice, tomado de un estudio reciente sobre la satisfacción laboral. Una de las variables claves era *depresión laboral*, la cual se medía con un índice compuesto de los siguientes cuatro reactivos que preguntaban a los trabajadores qué pensaban de ellos mismos y de su empleo:

- “Me siento triste y descorazonado.”
- “Estoy cansado sin razón.”
- “Me siento inquieto y no me mantengo tranquilo.”
- “Estoy más irritable que de costumbre.”

Los investigadores Amy Wharton y James Baron indicaron que "los reactivos se codificaron así: 4 = a menudo, 3 = a veces, 2 = raramente, 1 = nunca" (Wharton y Baron, 1987:578). Después explicaron como midieron otras variables que examinaron en el estudio:

La autoestima laboral se basó en cuatro reactivos que preguntaban a los entrevistados como se veían en su trabajo: felices/tristes; exitosos/fracasados; importantes/no importantes, esforzados al máximo/no esforzados al máximo. Cada reactivo calificaba de 1 a 7, el punto 1 indica una imagen personal de no ser feliz, exitoso ni importante, ni de esforzarse al máximo.

Al repasar la bibliografía de la investigación social, encontrará muchos ejemplos parecidos de índices acumulados para medir variables.

Manejo de los datos faltantes

Cualquiera que sea su método de recopilación de datos, enfrentará a menudo el problema de datos faltantes. Por ejemplo, en un análisis de contenidos de las tendencias políticas de los periódicos podría descubrir que uno de los diarios nunca ha adoptado una posición editorial sobre uno de los temas estudiados; digamos, que nunca se ha pronunciado sobre la Organización de las Naciones Unidas. En un diseño experimental que implica varias pruebas posteriores de los sujetos, algunos pueden no participar en varias sesiones. Prácticamente en todas las encuestas algunos interrogados no responden todas las preguntas (o escogen la opción "no sé"). Los datos faltantes plantean problemas en todas las etapas del análisis, pero es especialmente problemático cuando faltan en la elaboración de índices. Sin embargo, hay varios métodos para enfrentar estas dificultades.

Primero, si hay relativamente pocos casos con datos faltantes, puede excluirlos del índice y del análisis (tomé esa decisión en el ejemplo de la escuela de medicina). En esta situación, uno se preocupa principalmente por si los datos disponibles son suficientes para el análisis y por si la exclusión introduce una muestra sesgada cuando se utilizan los datos para el análisis. Puede evitar esto si la probabilidad mediante una comparación, en

otras variables importantes de los incluidos y los excluidos del índice.

Segundo, a veces uno tiene fundamentos para considerar los datos faltantes como una de las respuestas. Por ejemplo, si en un cuestionario se pidió a los encuestados que indicaran su participación en varias actividades marcando "sí" o "no" en cada una, muchos habrán señalado "sí" en algunas y dejado el resto en blanco. En tal caso, uno puede decidir que la falta de respuesta significa "no" y calificar los datos faltantes como si los entrevistados hubieran marcado el espacio respectivo.

Tercero, un análisis cuidadoso de los datos faltantes puede conducir a una interpretación de su significado. Por ejemplo, al elaborar una medida del conservadurismo político, podría descubrir que los entrevistados que no respondieron cierta pregunta eran en general tan conservadores en los otros reactivos como los que dieron una respuesta conservadora. En otro ejemplo, un estudio reciente de las creencias religiosas descubrió que quienes contestaban "no sé" acerca de determinada creencia eran casi idénticos a los "descreídos" en sus respuestas a otras creencias. (*Nota:* No tome estos ejemplos como reglas empíricas en sus propios estudios, sino como medios generales sugeridos para que analice sus datos.) Cuando del análisis de los datos faltantes se coligen estas interpretaciones, se puede tomar la decisión de calificar tales casos como corresponda.

Hay muchas otras formas de manejar este problema. Si un reactivo tiene varios valores posibles, podría asignar el peso central a los casos con datos faltantes; por ejemplo, asignar 2 si los valores son 0, 1, 2, 3 y 4. En las variables continuas, como la edad, también podría asignar la media a los casos faltantes. Asimismo, estos datos pueden suplirse asignando cifras al azar. Todas estas soluciones son conservadoras, en el sentido de que van en contra de cualesquiera relaciones que pretenda encontrar.

Si usted crea un índice de varios reactivos, a veces puede manejar los datos faltantes utilizando proporciones basadas en lo observado. Supongamos que su índice se compone de seis indicadores y que usted tiene sólo cuatro observaciones de cierto sujeto. Si el sujeto tiene 4 puntos de 4 posibles, podría darle una calificación en el índice de 4 y en otros 2 puntos (la mitad de los puntos

posibles en los cuatro reactivos), le asignaría una calificación de 3 (la mitad de los puntos posibles en seis observaciones).

La decisión de emplear un método depende a tal grado de la situación de investigación que no puedo sugerirle razonablemente un único método "óptimo" ni clasificar los que hemos estudiado. Excluir todos los casos con datos faltantes puede sesgar la representatividad de los hallazgos, pero incluirlos asignando puntuaciones a tales datos puede también influir en la naturaleza de los resultados. El camino mejor y más seguro es elaborar el índice con varios métodos para ver si se obtienen los mismos hallazgos. Comoquiera que sea, el objetivo final del análisis es entender los datos.

Validación de los índices

Hasta este momento hemos estudiado todas las etapas de la selección y la puntuación de los reactivos que dan por resultado un índice combinado que pretende medir alguna variable. Si se realizan cuidadosamente las etapas anteriores, aumenta la probabilidad de que el índice mida en realidad la variable. Sin embargo, para verificar lo anterior es preciso *validar* el índice. En la lógica básica de la validación, asumimos que el índice combinado da la medida de alguna variable; es decir, que las puntuaciones sucesivas del índice ordenan los casos de acuerdo con tal variable. Un índice del conservadurismo político ordena a las personas según su conservadurismo relativo. Si el índice consigue hacerlo, las personas que califica como relativamente conservadoras aparecerán como tales en todas las otras indicaciones de las tendencias políticas, como los reactivos de un cuestionario. Hay varios métodos para validar un índice combinado.

Análisis de reactivos La primera etapa para validar un índice es una **validación interna**, llamada *análisis de reactivos*. En este análisis se examina el grado al que el índice combinado se relaciona (o predice las respuestas) con los reactivos que comprende. Veamos un ejemplo de esta etapa.

En el índice de las orientaciones científicas de los catedráticos de la escuela de medicina, las puntuaciones iban de 0 (más interesado en la atención del paciente) a 3 (más interesado en la investiga-

ción). Ahora consideremos uno de los reactivos del índice: si los entrevistados querían profundizar sus conocimientos en el manejo integral del paciente o bien en el campo de los mecanismos básicos. Consideramos a los últimos como de inclinación más científica que los primeros. La siguiente tabla vacía muestra la manera en que examinaríamos la relación entre el índice y el reactivo.

	Índice de inclinaciones científicas			
	0	1	2	3
Porcentaje que manifestó más interés en los mecanismos básicos	??	??	??	??

Si dedica un minuto a reflexionar en la tabla, verá que ya conocemos las cifras que van en dos de las celdas. Para obtener una calificación de 3 en el índice, los encuestados tenían que responder "mecanismos básicos" a esta pregunta y dar también respuestas "científicas" en los otros dos reactivos. Así, 100 por ciento de los que tienen 3 en el índice respondieron "mecanismos básicos". Del mismo modo, todos los que recibieron 0 tuvieron que contestar el reactivo con "manejo integral del paciente". Así, cero por ciento de estos entrevistados dijo "mecanismos básicos". Veamos el aspecto de la tabla con la información que ya tenemos.

	Índice de inclinaciones científicas			
	0	1	2	3
Porcentaje que manifestó más interés en los mecanismos básicos	0	??	??	100

Si el reactivo es una buena muestra del índice general, esperaríamos llenar las celdas 1 y 2 con una progresión entre cero y 100 por ciento. Más de los que obtuvieron 2 deben haber escogido "mecanismos básicos" que de los que recibieron 1. Sin embargo, la forma de elaborar el índice no lo garantiza: se trata de una pregunta empírica, una

que respondemos con un análisis de reactivos. Así resulta el análisis de este reactivo en particular.

	Índice de inclinaciones científicas			
	0	1	2	3
Porcentaje que manifestó más interés en los mecanismos básicos	0	16	91	100

Como vemos, de acuerdo con nuestra suposición de que quienes obtuvieron 2 tienen una orientación más científica que quienes recibieron 1, descubrimos que un porcentaje mayor de los que tienen 2 (91 por ciento) respondió "mecanismos básicos" que de los que tienen 1 (16 por ciento).

El análisis de reactivos de los otros dos componentes del índice arrojaría resultados similares, como se muestra a continuación.

	Índice de inclinaciones científicas			
	0	1	2	3
Porcentaje que manifestó más interés en los mecanismos básicos	0	4	14	100
Porcentaje que dijo que prefería leer sobre las razones	0	80	97	100

Así, los tres reactivos parecen componentes adecuados del índice. Cada uno parece reflejar la misma cualidad que el índice en conjunto mide.

En un índice complicado con muchos reactivos, esta etapa ofrece una prueba conveniente de la contribución independiente de cada uno. Si resulta que cierto reactivo se relaciona poco con el índice, puede suponerse que otros reactivos cancelan su aportación. Si ese reactivo no añade nada a la fuerza del índice, lo mejor es excluirlo.

El análisis de reactivos es una importante primera prueba de la validez del índice, pero apenas es suficiente. Si el índice mide adecuadamente determinada variable, debe predecir otras indicaciones de dicha variable. Para probarlo, debemos acudir a los reactivos que no se incluyeron en el índice.

Validación externa La gente que califica como conservadora en un índice debe aparecer conservadora en sus respuestas a otros reactivos del cuestionario. Desde luego, hablamos de conservadurismo relativo, pues no podemos dar una definición final absoluta de lo que constituye el conservadurismo. Sin embargo, aquellos entrevistados que calificaron como los más conservadores en el índice deben ser los más conservadores al responder otras preguntas. Quienes calificaron como los menos conservadores, deben ser los menos conservadores en otros reactivos. En efecto, el ordenamiento de los grupos de interrogados en el índice debe predecir el orden con que estos grupos contesten otras preguntas sobre tendencias políticas.

En nuestro ejemplo del índice de tendencias científicas, varias preguntas del cuestionario ofrecen la posibilidad de tal **validación externa**. La tabla 7.1 presenta algunos reactivos que dan varias lecciones sobre la validación de índices. Primero, observamos que el índice predice bien las respuestas a los reactivos de validación en el sentido de que el orden de las respuestas científicas en los cuatro grupos es el mismo que en el propio índice. Al mismo tiempo, cada reactivo da una *descripción* diferente de las tendencias científicas generales. Por ejemplo, el último reactivo de validación indica que la gran mayoría de *todos* los catedráticos realizaron alguna investigación durante el año anterior. Si éste fuera el único indicador de la orientación científica, concluiríamos que casi todos los catedráticos son científicos. No obstante, quienes calificaron como más científicos en el índice tienden más a la investigación que quienes calificaron como relativamente menos científicos. El tercer reactivo de validación brinda un cuadro descriptivo distinto: sólo una minoría de los catedráticos preferiría que sus obligaciones se limitaran en exclusiva a la investigación, y los porcentajes de los que dan esta respuesta corresponden a las calificaciones asignadas en el índice.

Comparación entre índices deficientes e indicadores deficientes En algún momento, todos los que elaboran un índice deben enfrentar la aparente incapacidad de los reactivos externos para validarlo. Si el análisis de los reactivos internos muestra una relación inconstante entre los reactivos incluidos en el índice y el propio índice, hay en éste algo mal.

Tabla 7.1

Validación del índice de orientación científica

	Índice de orientación científica			
	Bajo 0	1	2	Alto 3
Porcentaje interesado en asistir a conferencias científicas en la escuela de medicina	34	42	46	65
Porcentaje que dijo que los catedráticos deberían tener experiencia como investigadores médicos	43	60	65	89
Porcentaje que preferiría que sus obligaciones consistieran únicamente en actividades de investigación	0	8	32	66
Porcentaje que realizó alguna investigación durante el año académico anterior	61	76	94	99

Pero si el índice no predice bien los reactivos de validación externa, la conclusión que se extrae es más ambigua. Se debe elegir entre dos posibilidades: (1) el índice no mide adecuadamente la variable en estudio, o (2) los reactivos de validación no miden adecuadamente la variable y por ende no son prueba suficiente del índice.

El investigador que ha trabajado mucho y a conciencia en la elaboración de un índice encontrará atractiva la segunda conclusión. Habitualmente, sentirá que ha incluido en el índice los mejores indicadores de la variable; por tanto, los reactivos de validación son indicadores de segunda mano. Sin embargo, debe usted aceptar que se pretende que el índice sea una medición poderosa de la variable; así, debe relacionarse con todo reactivo que toque la variable aunque sea en forma superficial.

Cuando falla la validación externa, debe examinar de nuevo el índice antes de decidir que los reactivos de validación son insuficientes. Una manera de hacerlo es examinar las relaciones entre los reactivos de validación y los reactivos incluidos en el índice. Si descubre que algunos de estos últimos se relacionan con los de validación y otros no, habrá aumentado su comprensión del índice tal como lo preparó inicialmente.

No hay una receta para solucionar este dilema; es una agonía a la que tienen que aprender a sobrevivir los investigadores serios. En última instancia, la sabiduría de su decisión de aceptar un índice estará determinada por su utilidad en los análisis posteriores. Quizá al principio piense que el índice era bueno y que lo que falla son los reactivos de validación, pero más tarde descubra que la variable (tal como la mide el índice) no se relaciona con otras variables en la forma que esperaba. Entonces, tal vez tenga que elaborar un nuevo índice.

Elaboración de índices del estatus de las mujeres

Hemos hablado de la elaboración de índices sobre todo en el contexto de la investigación mediante encuestas, pero otras clases de investigación también se prestan a estas mediciones compuestas. Por ejemplo, cuando la Organización de las Naciones Unidas (1995) se propuso examinar el estatus de las mujeres en el mundo, decidió crear dos índices para reflejar dos dimensiones.

El Índice de Desarrollo por Género (IDG) comparó mujeres y hombres en tres indicadores: esperanza de vida, educación e ingreso. Se acostumbra emplear estos indicadores para supervisar el estatus de las mujeres en el mundo. Los países escandinavos, Noruega, Suecia, Finlandia y Dinamarca, ostentan las más altas puntuaciones en estas mediciones.

El segundo índice, la Medición del Poder por Género (MPG), se orienta a las cuestiones del poder y comprende tres indicadores:

- La proporción de escaños parlamentarios que ocupan las mujeres
- La proporción de puestos administrativos, directivos, profesionales y técnicos que ocupan las mujeres
- Una medida del acceso al trabajo y los ingresos

Aquí también los países escandinavos tienen una clasificación elevada, pero se les unen Canadá, Nueva Zelanda, Holanda, Estados Unidos y Austria. Tener dos mediciones de la igualdad de los sexos permitió a los investigadores hacer distinciones más precisas. Por ejemplo, en varios países, en particular Grecia, Francia y Japón, las mujeres sa-

lieron relativamente bien en el IDG, pero bastante mal en la MPG; así, aunque tienen ventajas en términos de ingresos, educación y esperanza de vida, aún se les niega el acceso al poder. Y mientras que las puntuaciones en el IDG resultaron más elevadas en las naciones ricas que en las pobres, las puntuaciones en la MPG mostraron que el poder de las mujeres dependía menos de la riqueza nacional, pues al respecto muchos países en desarrollo sobrepasaron a algunas naciones industrializadas.

Al examinar varias dimensiones de las variables de su estudio, los investigadores de la ONU también descubrieron un aspecto de los ingresos de las mujeres que en general pasa inadvertido. Population Communications International (1996:1) resume bien el descubrimiento:

Cada año, las mujeres hacen una contribución invisible de 11000 billones de dólares a la economía mundial —dice el informe de la UNDP— en trabajo no asalariado y trabajo subpagado según los precios de mercado prevaletentes. Esta “subestimación” del trabajo de las mujeres no sólo mina su poder de compra, dice el HDR de 1995, sino que también reduce su ya de por sí bajo estatus social y mengua su capacidad de tener propiedades y aprovechar créditos. Mahbub ul Haq, el autor principal del informe, dice que “si el trabajo de las mujeres se reflejara con precisión en las estadísticas nacionales, saltaría en pedazos el mito de que los hombres son los principales sostenes del mundo”. El informe de la UNDP encuentra que las mujeres trabajan más horas que los hombres en casi todos los países, en deberes tanto pagados como no pagados. En los países en desarrollo, las mujeres realizan aproximadamente 53 por ciento de todo el trabajo y dedican dos tercios de su tiempo laboral a actividades no remuneradas. En los países industrializados, las mujeres hacen en promedio 51 por ciento del trabajo y como sus contrapartes en el mundo en desarrollo desempeñan alrededor de dos tercios de sus labores sin paga. Los hombres de los países industrializados son compensados por dos tercios de su trabajo.

Ahora dirigiremos nuestra atención de la creación de índices acumulados al examen de las técnicas de construcción de escalas. Aunque hay muchos métodos para construir escalas, limitaremos nuestra exposición fundamentalmente a cuatro —las

escalas de Bogardo, Thurstone, Likert y Guttman— y a un estudio del diferencial semántico.

Construcción de escalas

Los índices adecuados ofrecen una disposición ordinal de los casos de determinada variable. Todos los índices se basan en este supuesto: un senador que vota por siete proyectos de ley conservadores se considera más conservador que uno que sólo votó por cuatro. Ahora bien, los índices pueden dejar de tomar en cuenta que no todos los indicadores de una variable tienen la misma importancia o fuerza. El primer senador podría haber votado por siete proyectos moderadamente conservadores, mientras que el segundo habría votado en favor de cuatro muy conservadores (quizá pensó que los otros siete proyectos eran demasiado liberales y votó en contra).

Las escalas garantizan mejor la ordinalidad al considerar las *estructuras* entre los indicadores. Los reactivos de una medición combinada pueden tener diferentes *intensidades* en términos de la variable. Los cuatro procedimientos para construir escalas que describiremos ilustrarán la variedad de las técnicas disponibles.

Escala de distancia social de Bogardo

Supongamos que a usted le interesa conocer el grado al que nuestros compatriotas están dispuestos a asociarse, digamos, con los albaneses. Podría formular las siguientes preguntas:

1. ¿Está dispuesto a permitir que vivan albaneses en el país?
2. ¿Está dispuesto a permitir que vivan albaneses en su ciudad?
3. ¿Está dispuesto a permitir que vivan albaneses en su barrio?
4. ¿Está dispuesto a permitir que vivan albaneses en la casa de junto?
5. ¿Permitiría que su hijo, o hija, se casara con una albanesa, o albanés?

Observe que las preguntas aumentan en función de la cercanía del contacto que los encuestados quieren o no tener con los albaneses. A partir

del interés original de medir la disposición a asociarse con los albaneses, hemos formulado varias preguntas que indican diferentes grados de intensidad en esta variable. Esta clase de reactivos constituyen una **escala de distancia social de Bogardo**.

Las manifiestas diferencias de intensidad revelan una estructura entre los reactivos. Es de suponer que si una persona está dispuesta a aceptar alguna asociación, aceptará también todas las que la preceden en la lista. Por ejemplo, quien esté dispuesto a permitir que los albaneses vivan en su barrio, de seguro los aceptará en su ciudad y el país, pero puede o no estar dispuesto a acogerlos como vecinos o parientes. Así, ésta es la estructura lógica de intensidad inherente entre los reactivos.

Empíricamente, uno esperaría encontrar que son más las personas que aceptan compartir la ciudadanía que las que aceptan el matrimonio. En este sentido, hablamos de "reactivos fáciles" (por ejemplo, la residencia en el país) y de "reactivos difíciles" (por ejemplo, la unión matrimonial). Más personas están de acuerdo con los reactivos fáciles que con los difíciles. Con algunas excepciones inevitables, la lógica exige que una vez que una persona ha rechazado una relación presente en la lista, rechazará también las que le siguen, más difíciles.

La escala de distancia social de Bogardo ilustra la importante economía de las escalas como dispositivos de reducción de datos. Al conocer *el número* de relaciones con los albaneses que aceptará un entrevistado, sabemos *cuáles* aceptó. Así, una sola cifra resume con exactitud los datos de cinco o seis reactivos sin pérdida de información.

Escalas de Thurstone

A menudo, la estructura inherente de la escala de distancia social de Bogardo no es apropiada para las variables que se miden. De hecho, rara vez es obvia tal estructura lógica entre varios indicadores. Las **escalas de Thurstone** son un intento por crear un formato para generar grupos de indicadores de una variable que tengan por lo menos una estructura *empírica*. Uno de los formatos básicos es el de "intervalos de apariencia igual".

Se da a un grupo de jueces, digamos, 100 reactivos que se supone son indicadores de una variable. A cada uno se le pide que estime qué tan buen

indicador de la variable es cada reactivo y que le asigne una calificación entre un rango de, por ejemplo, 1 a 13. Si la variable fuera *prejuicios*, se pediría a los jueces que asignaran la calificación de 1 al indicador más débil de esta variable, 13 al más fuerte y puntuaciones intermedias a los que cayeran entre uno y otro extremos.

Cuando los jueces terminan su tarea, el investigador examina las calificaciones que asignaron a cada reactivo para determinar en cuáles estuvieron más de acuerdo. Aquellos reactivos en los que los jueces no coincidieron por márgenes amplios se descartarían por ambiguos. De los reactivos que produjeron un asentimiento general en las calificaciones, se escogerían uno o más para representar cada punto en la escala del 1 al 13.

Los reactivos seleccionados de esta manera se incluirían en el cuestionario de una encuesta. Se esperaría que los encuestados que resultaran prejuiciosos en los reactivos con una fuerza de 5, se mostrarían también prejuiciosos en los reactivos menos fuertes; si algunos interrogados no parecerían prejuiciosos en los reactivos con una fuerza de 6, se esperaría que tampoco resultarían prejuiciosos en aquellos con una fuerza mayor.

Si los reactivos de una escala de Thurstone se preparan y califican bien, se repetirían la economía y la eficacia de la reducción de datos inherentes a la escala de distancia social de Bogardo. Se podría asignar a cada entrevistado una sola calificación (la fuerza del reactivo más difícil que hayan aceptado) que representaría adecuadamente las respuestas a varios reactivos del cuestionario. Y, al igual que en la escala de Bogardo, el entrevistado que calificara 6 se consideraría más prejuicioso que el que obtuviera 5 o menos.

Las escalas de Thurstone no se usan mucho en la investigación actual, sobre todo por el enorme gasto de energía y de tiempo necesario para que 10 o 15 jueces califiquen los reactivos. Debido a que la calidad de sus evaluaciones dependerá de su experiencia con la variable que se estudia, se requerirían investigadores profesionales. Más aún, los significados que comunican los reactivos para indicar una variable cambian con el tiempo. Así, un reactivo que tiene cierto valor en un momento podría tener uno muy distinto más adelante. Para que una escala de Thurstone sea eficaz, habría que actualizarla periódicamente.

Escalas de Likert

Quizá haya escuchado que llaman **escala de Likert** a los reactivos de un cuestionario que contienen categorías de respuesta del tipo “Muy de acuerdo”, “De acuerdo”, “En desacuerdo” y “Muy en desacuerdo”. Técnicamente, el nombre es inapropiado, aunque Rensis Likert creó este formato de preguntas tan usado.

El valor particular del formato es la *ordinalidad* inequívoca de las categorías de respuesta. Si se permitiera a los entrevistados proponer o elegir respuestas como “un poco de acuerdo”, “bastante de acuerdo”, “realmente de acuerdo”, etc., al investigador le resultaría imposible juzgar la fuerza relativa del asentimiento pretendido por los interrogados. El formato de Likert resuelve el problema.

Sin embargo, Likert tenía algo más en mente. Concibió un método con el cual este formato pudiera utilizarse para determinar la *intensidad relativa* de los reactivos. Como ejemplo sencillo, supongamos que queremos medir los prejuicios contra las mujeres. Para ello, conformamos un conjunto de 20 enunciados, cada uno de los cuales manifiesta esos prejuicios. Uno de los reactivos podría ser “las mujeres no manejan tan bien como los hombres”. Otro, “no se debería dejar votar a las mujeres”. La intensidad de estos dos reactivos le parecerá clara, y la técnica de construcción de escalas de Likert demostraría tal diferencia, además de que asentaría la intensidad de los otros 18 enunciados.

Supongamos que le pedimos a una muestra de personas que concuerde o no con los 20 enunciados. Limitarse a conceder un punto por cada uno de los indicadores de prejuicios contra las mujeres daría la posibilidad de calificaciones de índices de 0 a 20. La escala de Likert da un paso más y calcula la puntuación de *índice promedio* de todos los que coinciden con cada uno de los enunciados. Digamos que todos los que estuvieron de acuerdo en que las mujeres son peores conductoras que los hombres tuvieron una calificación de índice promedio de 1.5 (de 20 posibles). Los que estuvieron de acuerdo en que no se debería dejar votar a las mujeres tuvieron una calificación de índice promedio de, por ejemplo, 19.5, lo que indica el mayor grado de prejuicios reflejado en esa respuesta.

Como resultado de este análisis de reactivos, se podría calificar de nuevo a los entrevistados para formar una escala: 1.5 puntos por estar de acuer-

do en que las mujeres manejan mal, 19.5 puntos por decir que las mujeres no deberían votar y diversos puntos por otras respuestas que reflejen la manera en que se relacionan los reactivos con el índice sencillo inicial. Si los que *discrepan* con el enunciado “votaría por una mujer para presidente” tuvieron una calificación de índice promedio de 15, entonces la escala daría 15 puntos a las personas que no estuvieran de acuerdo con el enunciado.

En la práctica, las escalas de Likert se usan rara vez en la actualidad. No sé por qué; tal vez parezcan demasiado complicadas. Sin embargo, el formato de reactivos concebido por Likert es uno de los que más se emplean en el diseño contemporáneo de cuestionarios. Habitualmente, ahora se aprovecha en la creación de índices simples. Con, digamos, cinco categorías de respuesta, se asignarían de 0 a 4 o de 1 a 5 puntos, tomando en cuenta la dirección de los reactivos (por ejemplo, una calificación de 5 a “muy de acuerdo” por los reactivos positivos y a “muy en desacuerdo” por los negativos). Entonces, cada entrevistado recibiría una calificación general que represente la suma de las puntuaciones que obtuvo por las respuestas de los reactivos.

Diferencial semántico

Al igual que el formato de Likert, el **diferencial semántico** le pide a los entrevistados que elijan entre dos posiciones opuestas. Veamos cómo funciona.

Supongamos que usted evalúa la eficacia de una nueva lección de apreciación musical en función de la capacidad que demuestran los sujetos para apreciar la música. Como parte de su estudio, usted desea tocar algunas selecciones y pedirle a los sujetos que refieran sus impresiones sobre lo que oyen. Una buena forma de captar estas impresiones sería usar un formato de diferencial semántico.

Para empezar, usted debe determinar las *dimensiones* en las que los sujetos juzgarán las selecciones. Entonces, necesita encontrar dos términos *opuestos* que representen los extremos de cada dimensión. Supongamos que una dimensión que le interesa es simplemente si los sujetos disfrutaban la pieza o no. En este caso, los términos opuestos serían “agradable” y “desagradable”. Del mismo modo, tal vez quiera saber si consideran las selecciones “complicadas” o “simples”, “armoniosas” o “disonantes”, etcétera.

Figura 7.5
Diferencial semántico: impresiones de selecciones musicales

	Mucho	Un poco	Indiferente	Un poco	Mucho	
Agradable	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Desagradable
Simple	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Complicada
Disonante	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Armoniosa
Tradicional	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Moderna
			etc.			

Una vez determinadas las dimensiones pertinentes y encontrados los términos que representan los extremos de cada una, podría preparar una hoja de calificación de cada pieza para que los sujetos la llenen. La figura 7.5 muestra el aspecto que tendría.

En cada línea de la hoja, el sujeto indicaría sus impresiones sobre la pieza musical; por ejemplo, si fue agradable o desagradable y si lo fue "un poco" o "mucho". Para no crear un esquema de respuestas tendencioso en esos reactivos, es una buena idea variar la colocación de los términos que probablemente se relacionen. Por ejemplo, observe que "disonante" y "tradicional" están en el lado izquierdo de la hoja, en tanto que en el derecho se encuentran "armoniosa" y "moderna". Lo más probable es que las selecciones calificadas de "disonantes" también califiquen de "modernas", en oposición a "tradicionales".

Tanto el formato de Likert como el diferencial semántico tienen más rigor y estructura que otros formatos de preguntas. Como dijimos, estos formatos arrojan datos adecuados lo mismo para índices que para escalas.

Escalas de Guttman

En la actualidad, los investigadores emplean a menudo las escalas que ideó Louis Guttman. Como las escalas de Bogardo, Thurstone y Likert, las de Guttman se basan en el hecho de que algunos reactivos del estudio resultarán ser indicadores más extremos de la variable que otros. Un ejemplo bastará para ilustrar esta situación.

Como recordará, en el ejemplo anterior sobre la medición de las tendencias científicas entre los catedráticos de la escuela de medicina elaboré un in-

dice simple. Sin embargo, resulta que los tres reactivos del índice forman en esencia una **escala de Guttman**.

La construcción de una escala de Guttman comienza con algunas de las mismas etapas que inician la elaboración de un índice. Empezaríamos por examinar la validez aparente de los reactivos disponibles para el análisis. Entonces, examinaríamos sus relaciones bivariadas y quizá multivariadas. Sin embargo, en la construcción de la escala también buscaríamos indicadores relativamente "difíciles" y "fáciles" de la variable que estamos examinando.

Cuando hablamos de las actitudes en cuanto al derecho de las mujeres al aborto, citamos varias condiciones que influyen en las opiniones de las personas: si la mujer es casada, si su salud está en peligro, etc. Estas condiciones proveen una excelente ilustración de las escalas de Guttman.

Éstos son los porcentajes de las personas que en la muestra de la ESG de 1996 apoyaron el derecho de las mujeres a abortar, en tres condiciones diferentes:

La salud de la mujer está en grave peligro	92%
El embarazo es resultado de una violación	86%
La mujer no está casada	48%

La diferencia en las tres condiciones en los porcentajes del respaldo al derecho de abortar revela algo acerca de los *niveles* de apoyo que cada reactivo indica. Por ejemplo, si alguien favoreciera el derecho a abortar cuando está en grave peligro la vida de la madre, no sería un indicador muy fuerte de respaldo al aborto, porque casi todos estarían de acuerdo. Respalda el derecho a abortar de las mujeres solteras parece un indicador mucho más

Tabla 7.2

Escala del apoyo al derecho de abortar

	Salud de las mujeres	Resultado de violación	Mujer soltera	Número de casos
Tipos de escala	+	+	+	612
	+	+	-	448
	+	-	-	92
	-	-	-	79
				Total = 1231
Tipos mixtos	-	+	-	15
	+	-	+	5
	-	-	+	2
	-	+	+	5
				Total = 27

+ = en favor del derecho de la mujer a elegir; - = se opone al derecho de la mujer a elegir

fuerte de apoyo al aborto en general: menos de la mitad de la muestra adoptó esta posición.

Las escalas de Guttman se basan en la noción de que cualquier persona que dé un indicador fuerte de alguna variable también dará los indicadores débiles. En esta situación, supondríamos que cualquiera que respaldara el derecho a abortar de las solteras también lo favorecería en caso de violación o si la salud de la mujer está amenazada. La tabla 7.2 pone a prueba este supuesto presentando la cifra de entrevistados que dieron cada uno de los esquemas de respuestas posibles.

Los primeros cuatro esquemas de respuesta de la tabla componen lo que podríamos llamar los *tipos de la escala*: aquellos esquemas que forman una estructura escalar. Después de los entrevistados que apoyan el derecho al aborto en las tres condiciones (línea 1), vemos (línea 2) que aquellos con sólo dos respuestas favorables eligieron las más fáciles, y que quienes tienen sólo una respuesta de este tenor (línea 3) optaron por la más fácil de las tres (la salud de la mujer está en peligro). Por último, se encuentran algunos entrevistados que se oponen al aborto en las tres circunstancias (línea 4).

La segunda parte de la tabla presenta los esquemas de respuesta que infringen la estructura escalar de los reactivos. Las diferencias más radicales de la estructura escalar son los dos últimos esquemas: los que aceptan sólo el reactivo más difícil y los que rechazan sólo el más fácil.

La última columna de la tabla indica el número de entrevistados que dieron todos los esquemas de

respuesta. Es obvio de inmediato que la gran mayoría (98 por ciento) se ajusta a uno de los tipos de escala. Sin embargo, la presencia de tipos mixtos indica que los reactivos no forman una escala de Guttman *perfecta*.

En este punto debemos recordar que una de las principales funciones de la construcción de una escala es la reducción eficaz de los datos. Las escalas ofrecen una técnica para presentar los datos en forma resumida al tiempo que conservan tanta información original como sea posible.

Cuando conformamos un índice con los reactivos de la orientación científica asignamos a los entrevistados un punto por cada respuesta científica que dieron. Si calificáramos los mismos tres reactivos como una escala de Guttman, asignaríamos a algunos entrevistados puntuaciones que permitirían la reproducción más exacta de sus respuestas originales.

Los entrevistados que se ajustaran a los tipos de escala recibirían las mismas puntuaciones que les asignamos en la elaboración del índice. Las personas que dieron las tres respuestas favorables calificarían aún 3, quienes eligieron las respuestas favorables en los dos reactivos más fáciles y se opusieron al más difícil recibirían 2, etc. En cada uno de los tres tipos de escala, podríamos predecir todas las respuestas reales que dieron los entrevistados basados en sus calificaciones.

Pero los tipos mixtos de la tabla presentan un problema. El primer tipo mixto (- + -) recibió 1 punto en el índice para señalar sólo una respuesta

Tabla 7.3
Puntuaciones de índice y escala

	Esquema de respuesta	Número de casos	Puntuaciones de índice	Puntuaciones de escala*	Errores totales de escala
Tipos de escala	+++	612	3	3	0
	++-	448	2	2	0
	+--	92	1	1	0
	---	79	0	0	0
Tipos mixtos	-+-	15	1	2	15
	+ - +	5	2	3	5
	--+	2	1	0	2
	- + +	5	2	3	5
				Errores totales de escala = 27	

$$\begin{aligned} \text{Coeficiente de reproducibilidad} &= 1 - \frac{\text{número de errores}}{\text{número de conjeturas}} \\ &= 1 - \frac{27}{1258 \times 3} = \frac{3747}{3774} \\ &= .993 = 99.3\% \end{aligned}$$

*Esta tabla muestra un método común para calificar los tipos mixtos, pero usted debe tener presente que también se aplican otros métodos.

favorable. Ahora bien, si se asignara 1 como calificación en la escala, predeciríamos que los 15 entrevistados de este grupo eligieron sólo el reactivo más fácil (permitir el aborto cuando la vida de la madre está en peligro) y cometeríamos dos errores con cada sujeto. Por tanto, las calificaciones de las escalas se asignan con objeto de reducir al mínimo los errores que se cometerían al reconstruir las respuestas originales.

La tabla 7.3 ilustra las puntuaciones de índice y escala que se asignarían a los esquemas de respuesta de nuestro ejemplo. Observe que en los tipos mixtos se comete un error con cada entrevistado. Es lo menos que podemos esperar en un esquema de tipo mixto. Por ejemplo, en el primero nos equivocariamos al predecir una respuesta favorable en el reactivo más fácil de los 15 entrevistados del grupo, lo que sumaría un total de 15 errores.

El grado al que un conjunto de respuestas empíricas forman una escala de Guttman se determina por la precisión con que se pueden reconstruir las respuestas originales de las puntuaciones. De cada uno de los 1 258 entrevistados de este ejemplo, predeciríamos tres respuestas del cuestionario, un total de 3 774 predicciones. La tabla 7.3 in-

dica que cometeríamos 27 errores si usamos las puntuaciones asignadas en la escala. El porcentaje de predicciones correctas se denomina **coeficiente de reproducibilidad**: el porcentaje de las respuestas originales que se pueden *reproducir* conociendo las puntuaciones de la escala que las resumen. En este ejemplo, el coeficiente de reproducibilidad es 3 747/3 774, o 99.3 por ciento.

Salvo por el caso de la reproducibilidad perfecta (100 por ciento), no hay manera de saber si un conjunto de reactivos forman o no una escala de Guttman en ningún sentido absoluto. Prácticamente todos los conjuntos de estos reactivos se *aproximan* a una escala. Sin embargo, como lineamiento general, suelen tomarse como criterios al respecto los coeficientes de 90 o 95 por ciento. Si la reproducibilidad observada excede el nivel impuesto, bien puede uno decidir puntuar y emplear los reactivos como escala.

Desde luego, la decisión sobre los criterios a este respecto es arbitraria. Más aún, un grado elevado de reproducibilidad no asegura que la escala elaborada mida de hecho el concepto que se estudia, aunque sí incrementa la confianza en que todos los reactivos componentes midan lo mismo. Además, advierta que es más probable obtener un

coeficiente de reproducibilidad alto si se tienen pocos reactivos.

Debemos hacer un último comentario con respecto a las escalas de Guttman, a saber, que se basan en la estructura observada entre *los datos reales en estudio*. Éste es un punto importante que a menudo se entiende mal. No tiene sentido decir que un conjunto de reactivos de cuestionario (quizá los que preparó y utilizó un investigador anterior) constituyen una escala de Guttman. Todo lo que podemos decir es que forman una escala en el contexto de cierto grupo de datos que se van a analizar. Así, la capacidad de formar una escala es una cuestión empírica que depende de la muestra. Aunque un conjunto de reactivos formen una escala con una muestra de entrevistados, no hay ninguna garantía de que el mismo conjunto formará tal escala con otra muestra. En este sentido, los reactivos de un cuestionario nunca forman en y por sí mismos una escala, sino que ésta es producto de un conjunto de observaciones empíricas.

Tipologías

Concluimos este capítulo con una breve exposición de la elaboración y el análisis de tipologías. Recuerde que se elaboran índices y escalas para tener mediciones ordinales de las variables dadas. Tratamos de asignar puntuaciones de índice o escala a los casos para que indiquen un grado creciente de prejuicios, religiosidad, conservadurismo, etc. En tales casos, nos ocupamos de dimensiones únicas.

Sin embargo, muchas veces el investigador quiere resumir la intersección de dos o más variables y crear un conjunto de categorías o tipos, lo que llamamos una *tipología*. Por ejemplo, quizá usted desee examinar por separado las tendencias políticas de los periódicos en cuanto a los asuntos internos y la política exterior. La presentación cuatripartita de la tabla 7.4 describe esa tipología.

Los periódicos de la celda A son conservadores tanto en política exterior como interior; los de la celda B son liberales en ambas. Aquellos de las celdas B y C son conservadores en una y liberales en la otra.

Con frecuencia, se llega a una tipología en el curso de elaborar un índice o una escala. Resulta que los reactivos que uno piensa que representan

Tabla 7.4

Tipología política de los periódicos

		Política exterior	
		Conservador	Liberal
Política interna	Conservador	A	B
	Liberal	C	D

una variable en realidad representan dos. Tal vez queríamos elaborar un índice simple de las tendencias políticas de los periódicos, pero descubrimos empíricamente que había que mantener separadas la política exterior y la interior.

En cualquier caso, debe estar al tanto de una dificultad inherente al análisis tipológico. Probablemente no tendrá problemas cada vez que se toma una tipología como la **variable independiente**. En el ejemplo anterior, calcularíamos los porcentajes de los periódicos de cada celda que por lo común se adhieren a los candidatos demócratas; entonces, examinaríamos con facilidad los efectos de los asuntos externos e internos en los respaldos políticos.

Sin embargo, es extremadamente difícil analizar una tipología como **variable dependiente**. Si queremos descubrir por qué se disponen los periódicos en esas celdas de la tipología, estamos en problemas. Esto se hace obvio si consideramos las formas en que podríamos elaborar y leer las tablas. Por ejemplo, supongamos que queremos examinar los efectos del tamaño de la población en los sistemas políticos. Con una sola dimensión determinaríamos fácilmente los porcentajes de periódicos rurales y urbanos que calificaron como conservadores o liberales en el índice o la escala.

En cambio, con una tipología tendríamos que presentar la distribución de los periódicos urbanos de la muestra en los tipos A, B, C y D. En seguida habría que repetir el procedimiento con las publicaciones rurales de la muestra y comparar las dos distribuciones. Supongamos que 80 por ciento de los periódicos rurales calificaron como tipo A (conservadores en ambas dimensiones) en comparación con 30 por ciento de los urbanos. Además, digamos que solo 5 por ciento de los periódicos rurales calificaron como tipo B (conservadores sólo en los asuntos internos) en comparación con 40

por ciento de los urbanos. Sería incorrecto concluir de un examen del tipo B que los periódicos urbanos son más conservadores en los asuntos internos que los rurales, puesto que 85 por ciento de éstos tienen esta característica, en comparación con 70 por ciento de aquéllos. Esta relativa escasez de los periódicos rurales en el tipo B se debe a su concentración en el tipo A. Es evidente que sería muy difícil interpretar estos datos para cualquier otra cosa aparte de las descripciones.

En realidad, es más probable que uno examine estas dos dimensiones por separado, especialmente si la variable dependiente tiene más categorías de repuestas que el ejemplo dado.

No piense que hay que evitar siempre las tipologías en la investigación social; a menudo son el medio más apropiado para comprender los datos. Sin embargo, tenga cuidado con las dificultades especiales del uso de tipologías como variables dependientes.

Puntos principales

- Los indicadores simples de variables rara vez tienen una validez lo bastante clara para garantizar su uso.
- Las mediciones combinadas, como las escalas y los índices, resuelven el problema al incluir varios indicadores de la variable en una medición sumaria.
- Tanto las escalas como los índices pretenden ser mediciones ordinales de las variables, pero las escalas suelen cumplir este objetivo mejor que los índices.
- Los índices se basan en la mera acumulación de los indicadores de una variable.
- Las escalas aprovechan cualesquiera estructuras lógicas o empíricas de intensidad entre los indicadores de la variable.
- La validez aparente es el primer criterio que deben satisfacer los indicadores para incluirlos en una medición combinada; es decir, debe parecer (valor *aparente*) que proporcionan alguna medida de la variable.
- Si en realidad varios reactivos son indicadores de la misma variable, deben estar relacionados. Por ejemplo, si la frecuencia de la asistencia a la iglesia y la frecuencia con que reza una persona son indicadores de su religiosidad, entonces debe resultar que aquellos que van a menudo a la iglesia también rezan más que quienes asisten a la iglesia con menos frecuencia.
- Una vez elaborado el índice o la escala, hay que validarlos. La validación interna se refiere a las relaciones entre los reactivos incluidos en la medición combinada y la propia medición. La validación externa se refiere a las relaciones entre la medición combinada y otros indicadores de la variable (*no* incluidos en esa medición).
- La escala de distancia social de Bogardo es un medio para medir en qué grado una persona está dispuesta a asociarse con cierta clase de gente, como alguna minoría étnica. Se pide a los sujetos que indiquen si están o no dispuestos a aceptar varias clases de asociación. Las respuestas a estas preguntas se pueden resumir en una sola puntuación que represente la asociación más estrecha aceptable, porque quienes están dispuestos a aceptarla también estarían dispuestos a aceptar las más lejanas.
- Las escalas de Thurstone son una técnica para crear indicadores de variables que tengan una clara estructura de intensidad. Unos jueces determinan la intensidad de los indicadores.
- Las escalas de Likert son una técnica de medición basada en el uso de categorías de respuesta estandarizadas (por ejemplo, "muy de acuerdo", "de acuerdo", "en desacuerdo", "muy en desacuerdo") para varios reactivos del cuestionario. El formato de reactivos de Likert se puede emplear con provecho en la elaboración de índices o escalas.
- El diferencial semántico es un formato de pregunta que pide a los entrevistados que indiquen puntuaciones entre dos extremos, como "muy positivo" y "muy negativo".
- Quizá la técnica de construcción de escalas más popular en la actualidad sea la de Guttman, un método para descubrir y utilizar la estructura empírica de intensidad entre varios indicadores de la variable dada.
- Los coeficientes de reproducibilidad son una medida del grado al que las respuestas dadas a los reactivos incluidos en una escala se pueden reproducir a partir sólo de las puntuaciones de la escala.

- Las tipologías son mediciones combinadas nominales que se usan con frecuencia en la investigación social. Las tipologías se aprovechan con eficacia como variables independientes, pero es difícil interpretarlas cuando se toman como variables dependientes.

Preguntas y ejercicios de repaso

1. Explique con sus propias palabras la diferencia entre índice y escala.
2. Redacte tres reactivos de un cuestionario que midan las actitudes hacia la energía nuclear y que sea probable que formen una escala de Guttman.
3. Los economistas suelen emplear índices para medir variables económicas, como el costo de la vida. Visite la página en internet de la Oficina Estadounidense de Estadísticas Laborales (Bureau of Labor Statistics, <http://stats.bls.gov>) y busque la encuesta del índice de precios al consumidor. ¿Qué dimensiones del costo de la vida se incluyen en esta medición?
4. Encuentre un ejemplo de medida compuesta presentada en los medios de comunicación: periódicos, revistas, televisión o internet. Detalle los elementos que componen la medida.

Proyecto de continuidad

Cree tres indicadores de las actitudes sobre la igualdad sexual que representen una escala creciente. Indique cuál es el indicador más fuerte y cuál el más débil.

Lecturas adicionales

Anderson, Andy B., Alexander Basilevsky y Derek P. J. Hum, "Measurement: Theory and Techniques", en Peter H. Rossi, James D. Wright y Andy B. Anderson (comps.), *Handbook of Survey Research*, Nueva York, Academic

Press, 1983, pp. 231-287. La lógica de la medición analizada en el contexto de las mediciones combinadas.

Bobo, Lawrence, y Frederick C. Licari, "Education and Political Tolerance: Testing the Effects of Cognitive Sophistication on Target Group Effect", en *Public Opinion Quarterly* 53 (otoño de 1989): 285-308. Los autores emplean una variedad de técnicas para determinar la mejor forma de medir la tolerancia hacia diversos grupos de la sociedad.

Glock, Charles, Benjamin B. Ringer y Earl Babbie, *To Comfort and to Challenge: A Dilemma of the Contemporary Church*, Berkeley, University of California Press, 1967. Estudio empírico que ilustra las mediciones combinadas. Puesto que se capta mejor la elaboración de índices y escalas con ejemplos concretos, éste podría ser un buen estudio para examinar. Los autores utilizan varias mediciones combinadas y son relativamente claros en cuanto a los métodos que emplearon para elaborarlas.

Lazarsfeld, Paul, Ann Pasanella y Morris Rosenberg (comps.), *Continuities in the Language of Social Research*, Nueva York, Free Press, 1972, esp. sec. 1. Excelente colección de exposiciones conceptuales y ejemplos concretos. La elaboración de mediciones combinadas se presenta en el contexto del campo más general de la conceptualización y la medición.

McIver, John P., y Edward G. Carmines, *Unidimensional Scaling*, Newbury Park, Cal., Sage, 1981. Excelente medio de profundizar en las escalas de Thurstone, Likert y Guttman.

Miller, Delbert, *Handbook of Research Design and Social Measurement*, Newbury Park, Cal., Sage, 1991. Excelente compilación de escalas usadas con frecuencia y seminormalizadas. Los numerosos ejemplos de la parte 4 del libro de Miller pueden adaptarse directamente a los estudios, o por lo menos sugieren mediciones modificadas. Más aún, el estudio de varios de los ejemplos aumentará su comprensión de la lógica de las mediciones combinadas en general.