



Universidad de Buenos Aires
Facultad de Ciencias Sociales
Carrera de Sociología

**Documento de Cátedra 34: Traducción de Frankfort-Nachmias y
Nachmias (1992) “Research Methods in the Social Sciences”.
Capítulos 7, 9, 10, 11 y 18.**

**Teresa Almendros, Florencia Arancibia, Mercedes Krause, Ignacia
Perugorria y Gabriela S. Plotno**

**Cátedra de Metodología y Técnicas de la Investigación Social
Profesora Titular: Ruth Sautu**

Este Documento de Cátedra forma parte de una serie que tiene como propósito contribuir a la formación de alumnos de la Carrera de Sociología de la Universidad de Buenos Aires en cuestiones vinculadas al diseño y realización de investigaciones científicas en el campo de las ciencias sociales. Su contenido complementa los textos de metodología de lectura obligatoria y optativa incluidos en cada uno de los tres cursos que se dictan en la Carrera.

Estos documentos son material de uso interno y no pueden ser incorporados a ediciones impresas ni reproducidos comercialmente.

La Cátedra solicita a los usuarios de estos Documentos que citen a sus autores indicando las referencias completas, es decir: autores, fecha, título, número y tipo de documento (traducción, resumen, elaboración propia, etc.). En los casos en que el usuario utilice sólo parte del documento, haciendo referencia a algunos de los autores/obras originales allí incluidos, consignar que fue tomado de nuestro Documento de Cátedra. Por ejemplo:

E. O. Wrigth (1985), *Classes*, London: Verso, citado en Documento de Cátedra II.1., Plotno, G., Lederman, F. & Krause, M. (2007) “Escalas Ocupacionales”.

CAPÍTULO 7

Medición

En este capítulo, exploraremos la estructura de medición y sus varios niveles de medición. También discutiremos acerca de la validez y la confiabilidad.

Cuando los investigadores se deciden por un problema de investigación y comienzan a especificar las hipótesis a examinar, se ven inmediatamente confrontados con el problema de cómo diseñar el estudio y cómo medir las variables. En los capítulos 5 y 6, hemos discutido acerca del diseño de investigación. Este capítulo enfoca la medición, su naturaleza y estructura, los niveles de medición, y la validez y la confiabilidad de los instrumentos de medición. El punto central a reconocer sobre la medición es, en las clásicas palabras de Norbert Wiener, “las cosas no... corren a nuestro alrededor con sus medidas estampadas como la capacidad de un auto transportista: requieren cierta cantidad de investigación para descubrir cuáles son sus medidas.”¹ En algunos casos esta investigación tomará la forma de una búsqueda de una medida desarrollada y reportada en la literatura profesional; en otros casos, el investigador tiene que desarrollar medidas a través de las observaciones empíricas requeridas por el problema de investigación y el diseño de investigación. Los investigadores tendrán que proveer evidencias de que las medidas son válidas y confiables.

La naturaleza de la medición

La **medición** está fuertemente atada a la idea de definiciones operacionales discutidas en el capítulo 2. Las definiciones operacionales son procedimientos de medición que construyen el nivel teórico-conceptual a través del nivel empírico-observacional. Más específicamente, la medición es un procedimiento por el cual uno asigna numeraciones –números u otros símbolos- a las propiedades empíricas (variables) de acuerdo a reglas.² Supongamos que intentas comprar un auto nuevo. Habiendo encontrado que la diferencia de precio entre varios autos compactos es mínima, decides hacer la compra basándote en cuál modelo se acerca más a los siguientes requerimientos: diseño, operación económica, y servicio. Estas tres características varían. Por ejemplo, un modelo puede estar bien diseñado y ser económico, pero el servicio suministrado por la fábrica puede ser insatisfactorio. Entonces, decides ordenar cada una de las tres características a partir de cinco números: 10, 11, 12, 13, y 14. El número 10 indica total insatisfacción, y el número 14 indica el mayor grado de satisfacción. Los números 11, 12, y 13 indican diferentes grados de satisfacción hacia la característica examinada. Examinas cinco modelos. La Tabla 7.1 resume la evaluación de cada modelo de acuerdo a los tres criterios que se le otorgan. Después de examinar los resultados, decides comprar el auto C porque

¹ Norbert Wiener, “A New Theory of Measurement: A Study in the Logia of Mathematics,” *Proceedings of the London Mathematical Society*, 19 (1920): 181. Citado en *Research Methods: Issues and sights*, ed. Billy J. Franklin and Harold W. Osborne (Belmont, Calif.: Wadsworth, 1971), p. 118.

² S. S. Stevens, “Mathematics, Measurement and Psychophysics,” in *Handbook of Experimental Psychology*, ed. S. S. Stevens (New York: Willey, 1951), p. 8.

recibe el mayor resultado en las tres cuentas, indicando el mayor grado de satisfacción.

Esta es una instancia extremadamente simplificada de medición, pero comunica la idea básica expresada en la definición. Le asignas numeraciones a las propiedades de acuerdo a reglas. Las propiedades o variables, las numeraciones, y las reglas de asignación estaban contenidas en las instrucciones específicas. Las numeraciones, que son el producto terminado de la medición, pueden usarse para la comparación, la evaluación, y el establecimiento de relaciones entre varias propiedades. Por ejemplo, puedes computar medidas de relación entre diseño y economía o entre diseño y servicio.

Tabla 7.1
Orden de preferencia

	Diseño	Economía	Servicio
Auto A	10	11	10
Auto B	13	14	12
Auto C	14	14	14
Auto D	14	12	13
Auto E	10	12	14

Definiendo la medición

Necesitamos más clarificación acerca de los tres conceptos básicos usados para definir la medición –numeraciones, asignaciones, y reglas-. Una *numeración* es un símbolo de la forma de I, II, III, ..., o 1, 2, 3, ... Una numeración no tiene ningún significado cuantitativo hasta que uno se lo da. Las numeraciones pueden ser usadas para identificar fenómenos, objetos, o personas como también meses, licencias de conducir, calles, libros, variables, o jugadores de fútbol. Las numeraciones a las que se les da significado cuantitativo se convierten en números; esto permite el uso de técnicas matemáticas y estadísticas con fines descriptivos, explicativos, y predictivos. En otras palabras, los números permiten un análisis cuantitativo, el cual puede revelar nueva información acerca de los ítems estudiados.

En la definición de medición, el término *asignación* significa conexión. Las numeraciones o números se conectan con objetos o eventos. La Figura 7.1 ilustra la idea de conexión en la medición: en la mezcla de círculos y cuadrados, 1 es conectado a los círculos, y 2 es conectado a los cuadrados.

El tercer concepto usado para definir la medición es el de *reglas*. Una regla especifica el procedimiento por el cual las numeraciones o números son asignados a objetos o eventos. Una regla podría ser: "Asigne las numeraciones desde 10 hasta 15 a los sistemas políticos de acuerdo a cuán democráticos son. Si un sistema político es muy democrático, tendría que otorgársele el número 15. Si un sistema político no es para nada democrático, tendría que otorgársele el número 10. A los sistemas políticos entre estos dos límites, asígnele números entre estos dos límites." O suponiendo que un grupo está compuesto por tres demócratas y dos republicanos; use la siguiente regla de conexión: "Si un individuo es demócrata, asígnele un 1; si es republicano, asígnele un 2." La aplicación de esta regla está ilustrada en la Figura 7.2.

Figura 7.1
Asignación de conexiones

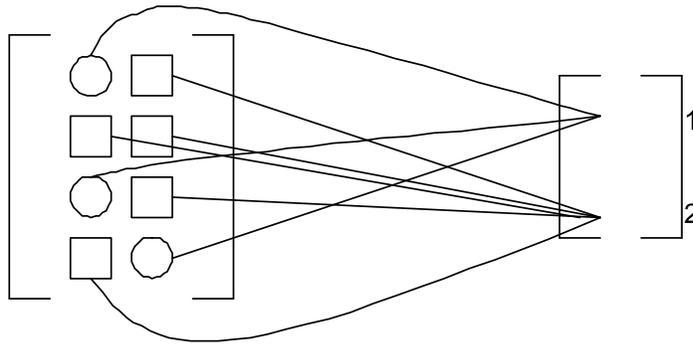
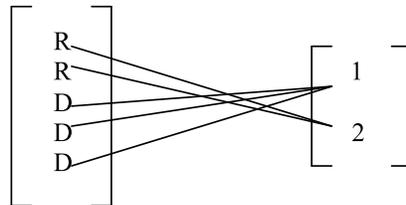


Figura 7.2
Asignación según la aplicación de una regla



Estructura de la medición

La medición, entonces, es la asignación de numeraciones o números a objetos, eventos, o variables de acuerdo a reglas. Las reglas son el componente más importante del procedimiento de medición porque determinan la calidad de la medición. Reglas pobres hacen que la medición pierda significado. La medición es insignificante cuando no está atada a la realidad, y la función de las reglas es atar el procedimiento de medición a la realidad. La medición significativa se alcanza sólo cuando el procedimiento de medición tiene una correspondencia empírica con la realidad. Por ejemplo, supongamos que alguien está midiendo la suavidad de tres objetos. Si el objeto A puede dañar al B y no viceversa, entonces B es más suave que A. Similarmente, si A puede dañar a B y B puede dañar a C, entonces A probablemente pueda dañar a C, y uno puede deducir que el objeto C es más suave que el objeto A. Estas son proposiciones observacionales, y luego de hacer algunos tests de daño puede asignársele números a cada uno de estos objetos indicando grados de suavidad. En este caso, el procedimiento de medición y el sistema de números son isomórficos a la realidad.

Isomorfismo significa “estructura similar o idéntica.” En medición, la pregunta crucial es si el sistema numérico es similar en estructura a la estructura de los conceptos que están siendo medidos. ¿Los dos son similares en algunos aspectos estructurales? Para el científico físico, el problema del isomorfismo es frecuentemente secundario porque la relación entre los conceptos observados y los números

asignados a las observaciones es casi directa. El científico social, en cambio, tiene que estar casi siempre alerta a este hecho:

Para que él pueda hacer ciertas operaciones con números que han sido asignados a observaciones, la estructura de este método de conectar números a las observaciones tiene que ser isomórfica a las estructuras numéricas que incluyen estas operaciones.³

Si dos sistemas son isomórficos, sus estructuras son las mismas en las relaciones y operaciones que permiten. Así si un investigador que asigna números a los objetos y luego los manipula, por ejemplo, sumándolos, está implicando que la estructura de esta medición es isomórfica a la estructura numérica aritmética.

Frecuentemente, los científicos sociales miden **indicadores** de conceptos. Los conceptos tales como democracia, motivación, hostilidad, y poder no pueden ser observados directamente; uno debe inferirlos de la medición de indicadores de los conceptos. Si las elecciones se efectúan regularmente en un sistema político, uno puede inferir que este es un indicador de democracia. Si uno alcanza un cierto resultado en un test de motivación, uno puede inferir algo acerca del nivel de motivación de esta persona. En estos ejemplos, algunos comportamientos identificables son indicadores de un concepto implícito. Muchas veces, se deben desarrollar múltiples indicadores para representar conceptos abstractos. Por ejemplo, democracia implica mucho más que elecciones. Elecciones justas, libertad de prensa, libertad de organización, y los derechos de las minorías son otros atributos esenciales. Los indicadores no deberían seleccionarse arbitrariamente. Ellos deben estar basados en los mundos teórico y empírico. Los indicadores usados para medir democracia en el ejemplo previo derivan de la teoría democrática y del comportamiento actual del sistema político. Los conceptos importantes de las ciencias sociales son multifacéticos y así requieren el uso de múltiples indicadores, donde cada uno refleja un aspecto distinto del concepto en cuestión. Aunque el procedimiento de medición de conceptos directamente observables sea idéntico a uno que mida a los indicadores de conceptos, las reglas del segundo serán más difíciles de formular porque será necesario un grado más amplio de inferencia. Así los indicadores son especificados por las definiciones operacionales; tras la observación de indicadores, los valores de los indicadores son sustituidos por numeraciones o números, y se efectúa el análisis cuantitativo. La estructura numérica que sustituye a los indicadores debe ser similar, en sus relaciones y operaciones, a la estructura de los indicadores; esto es, ambas deben ser isomórficas.

Niveles de medición

El requerimiento de isomorfismo entre sistemas numéricos y propiedades empíricas (o indicadores) nos lleva a una distinción alrededor de diferentes modos de medición, esto es, de diferentes niveles de medición. (El término *escalas de medición* es a veces usado en vez de *niveles de medición*. Una escala puede ser pensada como una herramienta de medición; un velocímetro es una escala, como lo es una regla o un cuestionario.) Las operaciones matemáticas y estadísticas posibles para un determinado grupo de números es dependiente del nivel de medición alcanzado. Aquí discutiremos acerca de cuatro niveles de medición –nominal, ordinal, intervalar, y razón- y los principios de las operaciones que están permitidas a cada nivel.

³ Sydney N. Siegel, *Nonparametric Statistics for the Behavioral Sciences* (New York: McGraw-Hill, 1988), p. 22.

Nivel nominal

El nivel de medición más bajo es el **nivel nominal**. En este nivel, los números u otros símbolos se usan para clasificar objetos u observaciones. Estos números o símbolos constituyen una escala nominal, o clasificatoria. Por el significado de los símbolos 1 y 2, por ejemplo, es posible clasificar una determinada población entre varones y mujeres, donde 1 representa a los varones y 2 a las mujeres. La misma población puede ser clasificada según religión; los cristianos pueden ser representados por la numeración 6, judíos por el 7, y musulmanes por el 8. En el primer caso, la población fue clasificada en dos categorías; en el segundo, en tres. Como regla, cuando un objeto puede ser clasificado en categorías que son exhaustivas (esto es, que incluyen a todos los objetos) y mutuamente exclusivas (esto es, sin ningún caso en más de una categoría) y cuando cada categoría es representada por un símbolo diferente, se trata de un nivel nominal de medición. Género, nacionalidad, etnia, religión, estado marital, lugar de residencia (por ejemplo, urbano o rural), e identificación partidaria son todas variables nominales.

Matemáticamente, la propiedad básica del nivel de medición nominal es que las propiedades de los objetos de una categoría son equivalentes entre sí, pero no a cualquier otra cosa en el mismo aspecto. Las propiedades lógicas de equivalencia son reflexividad, simetría, y transitividad. Reflexividad significa que cada objeto en una de las categorías es equivalente a sí mismo –por ejemplo, $a = a$ en la categoría “cristianos”. “Si $a = b$, entonces $b = a$ ” define la simetría, y “si $a = b$ y $b = c$, entonces $a = c$ ” expresa la transitividad. Estas tres propiedades lógicas son operativas alrededor de objetos de la misma categoría, pero no necesariamente entre categorías. Por ejemplo, estas relaciones se aplicarían a personas clasificadas en “cristianos” pero no entre “cristianos” y “judíos”.

En el nivel nominal, la clasificación de objetos puede ser llevada a cabo con cualquier grupo de símbolos. Los símbolos pueden también ser intercambiados sin alterar la información, si esto se hace consistentemente y completamente. De la misma manera, en el nivel nominal solamente están permitidas las estadísticas que permanecerían sin cambios tras una transformación tal. Estas incluyen el modo, las mediciones de variación cualitativa, y las mediciones de asociaciones apropiadas; estas estadísticas son presentadas en los capítulos 15 y 16.

Nivel ordinal

Muchas variables estudiadas por científicos sociales no sólo son clasificables sino que también exhiben alguna clase de relación. Relaciones típicas son “más grande”, “mejor”, “más deseado”, “más difícil”, y así sucesivamente. Semejantes relaciones pueden ser designadas por el símbolo $>$, que significa “superior a”. En referencia a propiedades particulares, $>$ puede ser usado para designar “es más grande que”, “es mejor que”, “es más deseado que”. Por ejemplo, una hipótesis puede ser que Francia es más democrática que la Unión Soviética pero menos democrática que Inglaterra o que los partidos políticos socialistas son menos dogmáticos que los comunistas pero más dogmáticos que los partidos religiosos. En general, si (en adición o equivalencia) la relación $>$ se mantiene para todos los pares de observaciones generando un ordenamiento completo de objetos, se trata de un **nivel ordinal** de medición. La relación equivalente se mantiene alrededor de casos del mismo orden, mientras que la relación $>$ se mantiene entre cualquier par de órdenes.

La relación $>$ es irreflexiva, asimétrica, y transitiva. La irreflexividad es una propiedad lógica en donde no es verdad que para cualquier a , $a > a$. La asimetría significa que si $a > b$, entonces $b \not> a$. La transitividad significa que si $a > b$ y $b > c$, entonces $a > c$. En otras palabras, si una variable como “conservadurismo” se mide en

el nivel ordinal, uno puede inferir que si una persona A es más conservadora que la persona B, y que si B es más conservadora que C, entonces A es más conservadora que C, y que la relación $>$ se mantiene incluyendo a todos los individuos del grupo.

Para ejemplificar la medición en el nivel ordinal, considere la siguiente práctica común de medición de actitudes. Las actitudes se miden a través de los significados de una serie de preguntas, con las respuestas alternativas ordenadas ascendente o descendentemente. Por ejemplo, una de las preguntas usadas para medir alienación política es “La gente como yo tiene mucha influencia en las decisiones del gobierno.” Se le pide al entrevistado que marque el número que representa su grado de acuerdo o desacuerdo con esta proposición. La correspondencia entre los números y las respuestas puede ser hecha como en la Tabla 7.2. Otras preguntas sobre la misma actitud son presentadas al entrevistado, quien puede entonces ser ordenado de acuerdo a sus respuestas a todas las preguntas. Suponiendo que un investigador emplea diez proposiciones en total, donde cada una permite cuatro respuestas alternativas: 1 para “fuertemente de acuerdo”, 2 para “de acuerdo”, 3 para “en desacuerdo”, y 4 para “fuertemente en desacuerdo”. El resultado más alto que puede ser alcanzado en este caso es de 40 (esto es, un resultado de 4 en cada una de los diez preguntas), y el nivel más bajo es 10. Para simplificar el asunto, asumimos que los entrevistados responden todas las preguntas. Un entrevistado cuyo resultado es de 40 será tenido en cuenta como el más alienado y será ordenado primero. Otro, cuyo resultado está cerca de 40 –digamos, 36- será ordenado segundo, y así sucesivamente con cada individuo del grupo. El proceso de ordenamiento termina cuando todos los entrevistados son ordenados según sus resultados en el cuestionario de alienación política. La Tabla 7.3 muestra los resultados hipotéticos y el ordenamiento de siete entrevistados. El examen de la tabla revela que el entrevistado S_6 es el más alienado y el entrevistado S_7 el menos alienado.

Tabla 7.2
Escala ordinal del ordenamiento

Orden	Valor
1	Fuertemente de acuerdo
2	De acuerdo
3	En desacuerdo
4	Fuertemente en desacuerdo

Tabla 7.3
Individuos ordenados según sus resultados en un test de alienación política

Entrevistado	Resultado	Orden
S_1	10	7
S_2	27	3
S_3	36	2
S_4	25	4
S_5	20	5
S_6	40	1
S_7	12	6

El nivel ordinal de medición es el único con una transformación monótona; esto es, cualquier transformación preservativa del orden no cambia la información obtenida. No importa qué números uno le asigne a un par de objetos o a una categoría de

objetos siempre y cuando uno sea consistente. Es una cuestión de conveniencia el hecho de usar números bajos para los órdenes más altos, aunque generalmente nos referimos a rendimientos superiores como “primera clase” y a rendimientos progresivamente inferiores como “segunda clase” y “tercera clase”. Además de actitudes, otras variables ordinales son clase social, grados escolares, rangos militares, posiciones jerárquicas en las organizaciones, y participación en partidos políticos.

Los números asignados a los objetos ordenados se llaman *valores ordinales*. Los valores ordinales se asignan a objetos de acuerdo a la siguiente regla: al objeto en un extremo (mayor o menor) se le asigna el 1; al siguiente en tamaño, 2; al tercero en tamaño, 3; y así sucesivamente hasta el otro extremo, al que se le asigna el último número de la serie. En el ejemplo de la Tabla 7.3, a S_6 se le asignó el 1, a S_3 se le asignó el 2, al S_2 se le asignó el 3, al S_4 se le asignó el 4, al S_5 se le asignó el 5, al S_7 se le asignó el 6, y al S_1 se le asignó el 7. Es importante remarcar que los números ordinales indican el orden y nada más. Los números no indican que los intervalos entre ellos sean iguales, tampoco indican cantidades absolutas. No se puede asumir que porque los números están igualmente espaciados, las propiedades que representan estén también igualmente espaciadas. Si dos entrevistados tienen los órdenes 7 y 5 y otros dos tienen los órdenes 4 y 2, uno no puede inferir que las diferencias entre los dos pares sea equivalente.

El nivel ordinal permite transformaciones que no cambien el orden de las propiedades. De la misma manera, las operaciones matemáticas y estadísticas que no alteran el orden de propiedades están también permitidas. Por ejemplo, una estadística que describe la tendencia central de los números ordinales es la media. La media no será afectada por los cambios en cualquiera de los números mayores o menores a ella siempre y cuando se mantenga el número de observaciones mayores y menores ordenadas. Otras estadísticas apropiadas para el nivel ordinal son rango, gamma, y tau-b, discutidas en los capítulos 15 y 16.

Nivel intervalar

Si, además de ordenar un grupo de observaciones en términos de la relación $>$, uno también conoce la distancia exacta entre cada una de las observaciones y esta distancia es constante, entonces se ha alcanzado un **nivel intervalar**. Además de decir que un objeto es mayor a otro, uno puede también especificar por cuántas unidades el primero es mayor al segundo. Por ejemplo, mediante la medición intervalar es posible decir no sólo que Sue gana más que Mike sino también que Sue gana, digamos, \$5.000 más que Mike. Para hacer estas comparaciones cuantitativas, uno debe tener una unidad de medición; cuando se establece una unidad de medición, se alcanza un nivel intervalar de medición. Ejemplos de variables medidas en el nivel intervalar son ingreso, coeficiente intelectual (IQ), resultados de examen, cantidad de votantes, y frecuencia de crímenes. Un nivel intervalar de medición, entonces, es caracterizado por una unidad de medición común y constante que asigna un número real a un par de objetos en el grupo ordenado. En esta clase de medición, la razón entre cualesquiera dos intervalos (distancia) es independiente de la unidad de medición.

La estructura de un nivel intervalar de medición es tal que las diferencias entre las observaciones son isomórficas con la estructura aritmética. Los números pueden ser asignados a las posiciones de los objetos de tal manera que varias operaciones aritméticas pueden ser significativas aplicándose a las diferencias entre estos números. Las siguientes propiedades formales caracterizan el nivel intervalar de mediciones:

1. Unicidad: si a y b representan números reales, entonces $a + b$ y $a \times b$ representan un número real y sólo uno.
2. Simetría: si $a = b$, entonces $b = a$.
3. Comunicación: si a y b denotan números reales, entonces $a + b = b + a$, y $ab = ba$.
4. Sustitución: si $a = b$ y $a + c = d$, entonces $b + c = d$; y si $a = b$ y $ac = d$, entonces $bc = d$.
5. Asociación: si a , b , y c representan números reales, entonces $(a + b) + c = a + (b + c)$, y $(ab)c = a(bc)$.

Cualquier cambio en los números asignados a las observaciones debe preservar no sólo el orden sino también sus diferencias relativas. En lenguaje más formal, el nivel intervalar de medición es el único con transformación lineal. Así la información obtenida en este nivel no se altera si cada número se multiplica por una constante positiva y entonces una constante se suma a este producto. Todas las descripciones e inferencias estadísticas son aplicables a los datos intervalares.

Tabla 7.4
Niveles de medición y sus propiedades características

Nivel	Equivalencia	Más grande que	Intervalo fijo	Cero natural
Nominal	Sí	No	No	No
Ordinal	Sí	Sí	No	No
Intervalar	Sí	Sí	Sí	No
Razón	Sí	Sí	Sí	Sí

Nivel razón

Las variables que tienen puntos cero naturales pueden ser medidas en el **nivel de medición razón**. Las variables tales como peso, tiempo, largo, y área tienen puntos cero naturales y son medidas en el nivel razón. En este nivel, la razón de cualesquiera dos números es independiente de la unidad de medición. Los niveles intervalar y razón son similares, y las reglas por las cuales los números son asignados son las mismas, con una excepción. Para un nivel de medición razón, aplicamos las operaciones y los números al monto total medido desde un punto cero absoluto; para un nivel de medición intervalar, aplicamos la operación a las diferencias desde un punto arbitrario. Un nivel de medición razón, más comúnmente encontrado en las ciencias físicas, es alcanzado sólo cuando estas cuatro relaciones son operacionalmente posibles de alcanzar: (1) equivalencia, (2) más grande que, (3) distancia entre cualesquiera dos intervalos conocida, y (4) un punto cero verdadero.⁴

Transformación de los datos

Las variables que pueden ser medidas en el nivel razón pueden también ser medidas en los niveles intervalar, ordinal, y nominal. Como regla, las propiedades que pueden ser medidas en un nivel superior pueden también ser medidas en niveles inferiores, pero no viceversa. Una variable como afiliación partidaria puede ser medida sólo en el nivel nominal. Las propiedades formales características de cada nivel de

⁴ Para más detalles sobre los niveles de medición, véase Siegel, *Nonparametric Statistics*, el cual informó gran cantidad de nuestra discusión.

medición son resumidas en la Tabla 7.4. Por ejemplo, mientras la propiedad de equivalencia existe en cada uno de los cuatro niveles, un cero natural caracteriza sólo al nivel razón.

Anteriormente mencionamos los tipos de operaciones matemáticas y estadísticas que son, en un sentido estricto, legítimas y permitidas en cada nivel. Algunos investigadores tienden a enfatizar esta cuestión. El problema, de todas maneras, es suficientemente significativo para garantizar unos pocos comentarios adicionales.

La matemática y la estadística son lenguajes sin contenido. Ellos tratan con números y no están interesados en el hecho de que los números representen al mundo empírico. Su ventaja más importante es la precisión y la posibilidad que les da a los investigadores de revelar información sobre un fenómeno que no puede ser revelado de otra manera. Una pregunta tal como “¿Hasta qué punto está relacionada una serie de variables?” puede ser significativa y precisamente contestada mediante la computación de mediciones de relaciones. Con números, puede desarrollarse cualquier tipo de operación estadística. Los científicos sociales se ocupan de fenómenos empíricos, y los números se usan principalmente para llegar a un mejor entendimiento de la relación entre estos fenómenos. El empleo de sistemas numéricos y estadísticas que no sean isomórficos a la estructura del fenómeno empírico es de poca ayuda para el avance del conocimiento.

Error de medición

Los procedimientos de medición asignan numeraciones, números, o resultados a las propiedades. Las diferencias en los resultados pueden ser atribuidas a dos fuentes. Una es el hecho de que las propiedades exhiben *diferencias reales* en los aspectos de la propiedad medida. La otra fuente de diferencias en los resultados es el hecho de que la medición en sí misma o el contexto en el que es llevada a cabo influencia a los resultados. En este caso, las mediciones están mostrando diferencias irreales. Las mediciones perfectas revelan sólo diferencias reales entre las propiedades. De todas formas, las mediciones raramente son perfectas y frecuentemente reflejan no sólo diferencias reales sino también diferencias artificiales producidas por el procedimiento de medición en sí mismo. Las diferencias que se deben a cualquier otra que no sea una diferencia real se llaman **errores de medición**. Ellos no se corresponden con las diferencias reales entre las propiedades sino que son diferencias debidas a la imperfección del procedimiento de medición.

Hay muchas fuentes de errores de medición comunes. Primero, los resultados obtenidos pueden estar relacionados a un atributo asociado. Por ejemplo, una pregunta que mida el desarrollo moral puede requerir un cierto nivel de inteligencia y conocimiento social para ser interpretada y contestada. Las respuestas de los individuos a esta pregunta reflejará no sólo las diferencias reales en el desarrollo moral sino también las diferencias en la inteligencia y el conocimiento social. Los errores de medición son efecto de los atributos asociados. En segundo lugar, los errores de medición pueden ser el resultado de las diferencias en las condiciones temporales, tales como salud o humor, que pueden afectar las respuestas de una persona al cuestionario o su comportamiento. En tercer lugar, los errores de medición pueden también ser el resultado de cómo diferentes personas interpretan el instrumento de medición de maneras diferentes. En cuarto lugar, las diferencias en el contexto contribuyen a los errores de medición. Por ejemplo, la edad, raza, y género de los entrevistadores influencia las respuestas de los entrevistados para la investigación. En quinto lugar, las diferencias en la administración del instrumento de medición (por ejemplo, poca luz, ruido, entrevistadores cansados) obtiene errores de medición. La última fuente principal de distorsiones son las diferencias en el procesamiento (por

ejemplo, diferentes códigos codifican respuestas similares de diferente manera) y análisis de datos.

Los errores que surgen de estas fuentes pueden ser sistemáticos o bien aleatorios. Los *errores sistemáticos* se producen siempre que el instrumento de medición es usado, y son constantes a través de casos y estudios. Ellos introducen una medida de invalidación a los descubrimientos. Los *errores aleatorios*, en contraste, afectan cada uso del instrumento de medición de una manera diferente. La validez y confiabilidad dependen de las técnicas para reducir errores de medición.

Validez

El problema de la validez surge porque la medición en las ciencias sociales es, con muy pocas excepciones, indirecta. Bajo estas circunstancias, los investigadores no están nunca completamente seguros de estar midiendo lo que ellos pretenden medir. La **validez** tiene que ver con la pregunta “¿Se está midiendo lo que se pretende medir?” Por ejemplo, ¿el recuento de votos mide el desarrollo político? ¿El acuerdo con la proposición “Este mundo es manejado por unas pocas personas en el poder, y no hay mucho que un hombre pueda hacer al respecto” es un indicador de la variable “alienación”? Para responder tales preguntas, el investigador debe proveer evidencia que sostenga que un instrumento de medición, de hecho, mide lo que aparentemente mide.

Se distinguen tres tipos básicos de validez, cada uno de los cuales concierne a un aspecto diferente de la situación de medición: validez de contenido, validez empírica, y validez de constructo. Cada una incluye varios tipos de evidencia y tiene valor especial bajo ciertas condiciones.

Validez de contenido

Hay dos variedades comunes de validez de contenido: validez aparente y validez de muestreo. La **validez aparente** descansa sobre la evaluación subjetiva del investigador acerca de la validez de un instrumento de medición. En la práctica, la validez aparente no se relaciona con la pregunta acerca de si un instrumento mide lo que el investigador quiere medir; más bien, se relaciona con hasta qué punto mide lo que aparentemente mide de acuerdo a la evaluación subjetiva del investigador. Por ejemplo, un investigador intenta medir la variable “liberalismo” mediante un cuestionario que consiste de diez proposiciones. Después de construir el cuestionario, el investigador revisa cada proposición para evaluar hasta qué punto está relacionada con “liberalismo”. Para esto, el investigador puede consultar a algunos especialistas (jueces). Si hay un acuerdo entre los jueces, el investigador propondrá que el cuestionario tiene validez aparente y que, consecuentemente, mide al “liberalismo”. Un desacuerdo entre los jueces dañaría la validez aparente de un instrumento de medición.

El mayor problema con la validez aparente es que no hay procedimientos replicables para evaluar al instrumento de medición, así cada uno debe descansar enteramente sobre las resoluciones subjetivas.

La preocupación principal de la **validez de muestreo** es si una determinada población de situaciones o comportamientos es adecuadamente ejemplificada por el instrumento de medición en cuestión; esto es, ¿el contenido del instrumento representa adecuadamente el contenido de la población de la propiedad medida? La asunción tácita de la validez de muestreo es que toda variable tiene un contenido de población consistente en un gran número de ítems (proposiciones, preguntas, o indicadores) y que un instrumento altamente válido constituye una muestra

representativa de estos ítems. En la práctica, los problemas surgen cuando se define a una población contenida, en un constructo teórico, no empírico. (Estos problemas se discuten en el capítulo 8, en el cual se presentan las técnicas de muestreo.) Estos problemas dañan la efectividad de la validez de muestreo como test de la validez de un instrumento. Sin embargo, la validez de muestreo tiene una importante función: requiere familiaridad con todos los ítems de la población contenida. La validez de muestreo es especialmente común en investigaciones exploratorias, donde los investigadores intentan construir instrumentos y emplearlos por primera vez. Una vez que el instrumento ha sido usado, su validez puede ser evaluada con otros tests.

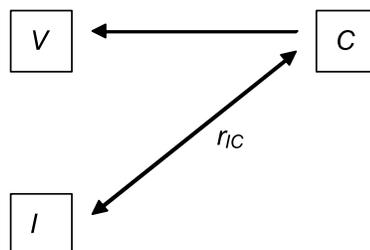
Validez empírica

La preocupación de la validez empírica tiene que ver con las relaciones entre un instrumento de medición y los resultados de medición. Se asume que si un instrumento de medición es válido, habría relaciones fuertes entre los resultados producidos por el instrumento y otras variables. La evidencia que sostenga la existencia de una relación es obtenida mediante las medidas de correlación apropiadas al nivel de medición. (Un coeficiente de correlación es un índice del grado de relación entre dos medidas; detalles pueden encontrarse en el capítulo 16.) De los tests varios designados para evaluar la validez empírica, la validez predictiva es la más ampliamente usada. Por esta razón, la detallaremos.

La **validez predictiva** es estimada mediante una predicción hacia una medida externa llamada *criterio* y el chequeo del instrumento de medición por medio de los resultados. En otras palabras, la validez predictiva es el coeficiente de correlación entre los resultados de una determinada medición y un criterio externo. Por ejemplo, uno puede validar un test de inteligencia obteniendo primero los resultados del test de un grupo tal como estudiantes universitarios y obteniendo luego los promedios que estos estudiantes alcanzaron en su primer año universitario. Entonces se computa un coeficiente de correlación entre los dos grupos de mediciones. Este coeficiente de correlación se llama *coeficiente de validación*. Otros criterios que podrían usarse para validar los tests de inteligencia son los resultados sociales en tests de adaptación y las numeraciones en el rendimiento.

El proceso por el cual la validez predictiva de un instrumento es evaluada está ilustrada en la Figura 7.3. Una variable (V) se mide por determinado instrumento (I), y el investigador desea evaluar la validez predictiva del instrumento. Para juzgar su validez predictiva, se usa un criterio (C) cuya validez es acordada. La medición obtenida por I está correlacionada con las mediciones obtenidas por C . El tamaño del coeficiente de validación (r_{IC}) mide la validez predictiva del instrumento.

Figura 7.3
Evaluación de la validez predictiva



Dos cuestiones generales deben ser consideradas cuando se usa el test de validez predictiva. Una se relaciona con la validez del criterio, y la otra concierne a las razones de uso de un instrumento de medición en vez del criterio en sí; por ejemplo, ¿por qué no medir los promedios directamente? Respecto al segundo punto, en algunos casos el criterio es técnicamente difícil o caro de usar, y en otros casos los investigadores tienen que medir una variable antes de hacer uso del criterio. Por ejemplo, las habilidades escolares de los estudiantes tienen que estar predeterminadas para su admisión en la universidad.

Respecto a la validez de criterio, se usan dos métodos comunes. Un método depende del acuerdo entre los investigadores acerca de un cierto criterio como válido para evaluar un instrumento de medición. El acuerdo depende de los tests de validez aparente y validez de muestreo. Un método de alguna forma diferente consiste en expresar la relación entre el instrumento y el criterio en términos del porcentaje de individuos (u otras unidades de análisis) que estaría correctamente clasificado por el instrumento de acuerdo a los grupos de pertenencia conocidos.⁵

Suponiendo que uno tiene que evaluar la validez de un instrumento designado para medir conservadurismo político. Si hay razones aparentemente teóricas como para argumentar que las personas de clase baja son más conservadoras que las personas de clase media, las dos clases pueden ser comparadas para chequear la validez predictiva. En este caso, la clase social sirve como criterio indirecto a la validez predictiva del instrumento. Si los descubrimientos revelan que las personas de clase baja son tan conservadoras como las personas de clase media, el instrumento carece de validez predictiva. Al revés, una correlación relativamente alta entre la clase social y el conservadurismo validaría el instrumento. De todas formas, la correlación alta es una condición necesaria pero no suficiente para comprobar la validez predictiva de un instrumento porque el criterio indirecto (la clase social) puede también estar relacionada con otras variables además del conservadurismo político (por ejemplo, educación). El instrumento puede medir otras variables en vez del conservadurismo político. Un criterio indirecto, entonces, es más útil para invalidar que para validar un instrumento de medición.

Validez de constructo

La **validez de constructo** implica relacionar un instrumento de medición con un marco teórico para determinar si el instrumento está atado a los conceptos y asunciones teóricas empleadas. Lee J. Cronbach, a partir de una de las primeras propuestas de validez de constructo, observó que “no importa lo que el test pregunte, lo que signifique un resultado psicológicamente, o lo que cause que una persona alcance determinado resultado en un test, se está preguntando qué conceptos son los apropiados para interpretar el rendimiento del test.”⁶ Las expectativas teóricas acerca de la variable medida llevan al investigador a postular varios tipos y grados de relaciones entre una variable particular y otras variables específicas. Para demostrar la validez de constructo de un instrumento de medición, un investigador debe mostrar que esas relaciones de verdad existen. Ilustraremos la utilidad de la validez de constructo a través de la famosa investigación de Milton Rokeach sobre dogmatismo.⁷

En la base del razonamiento teórico, Rokeach construyó un cuestionario sobre dogmatismo. El instrumento consistía de proposiciones asumidas para medir las mentalidades cerradas, una manera de pensar asociada a cualquier creencia o ideología sin importar su contenido. Rokeach argumentaba que las orientaciones

⁵ C. G. Helmstadter, *Research Concepts in Human Behavior* (Eaglewood Cliffs, N. J.: Prentice-Hall, 1970).

⁶ Lee J. Cronbach, *Essentials of Psychological Testing*, 4th ed. (New York: Basic Books, 1960).

⁷ Milton Rokeach, *The Open and the Closed Mind* (New York: Basic Books, 1960).

ideológicas de los individuos están relacionadas, mediante procesos, con sus personalidades y su comportamiento. En consecuencia, él predecía, entre otras cosas, que el dogmatismo está relacionado con las opiniones. Rokeach condujo numerosos estudios con el objeto de testear su teoría y la validez de constructo de los instrumentos de medición. En un estudio, usó la **técnica de los grupos-conocidos**. En este método, un instrumento de medición es administrado a grupos de personas con atributos conocidos, y se predice la dirección de las diferencias. Rokeach les pidió a los profesores de la universidad y a los estudiantes graduados que seleccionaran amigos que ellos creyeran con mentalidades cerradas o abiertas. El cuestionario de dogmatismo diferenciaba claramente los dos grupos. Este descubrimiento proveía evidencia de soporte a la validez de constructo de la medición del dogmatismo.

Cronbach y Meehl describen el proceso lógico de la validación del constructo de la siguiente manera: primero, se presenta una proposición de que el instrumento mide una cierta propiedad –digamos, propiedad A-; en segundo lugar, la proposición se inserta a la teoría presente en la propiedad A; en tercer lugar, trabajando a través de la teoría, uno predice otras propiedades que deberían estar relacionadas con el instrumento y las propiedades que no deberían exhibir ninguna relación con el instrumento; finalmente, uno colecta los datos que confirmen o rechacen empíricamente las relaciones predichas. Si se encuentran las relaciones anticipadas, el instrumento se considera válido. Si las predicciones fallan, hay tres posibilidades: (1) el instrumento no mide a la propiedad A, (2) el marco teórico que generó las predicciones es incorrecto, o (3) el diseño de investigación no pudo testear las predicciones apropiadamente. El investigador debe entonces tomar una decisión sobre cuál de estas tres condiciones ha ocurrido. Tal decisión está basada en una cuidadosa reconstrucción de cada uno de los cuatro pasos que constituyen el proceso de validación.⁸

Campbell y Fiske sugirieron otro método de validación de constructo que implica a las matrices de correlación.⁹ Esta es la concepción *discriminante-convergente* de validación, o la técnica de *matriz de multicualidad-multimétodo*. Este método deriva de la idea de que diferentes métodos que midan la misma propiedad deberían producir resultados similares, mientras que diferentes propiedades deberían producir resultados de medición diferentes sin importar el instrumento de medición. Operacionalmente, esto significa que los coeficientes de correlación entre los resultados de una determinada propiedad, medida por instrumentos diferentes, debería ser superior a las correlaciones entre propiedades diferentes, medidas por instrumentos similares. Por todo esto la evidencia de la validez de constructo de un instrumento debe hacer uso tanto del principio convergente –esto es, dos medidas de la misma propiedad deberían correlacionarse mutuamente aunque representen diferentes métodos- como del principio discriminatorio, el cual implica que dos medidas no deberían correlacionarse mutuamente si miden diferentes propiedades, aunque usen un instrumento similar. Así el proceso de validación llama a la computación de intercorrelaciones entre instrumentos de medición que representen por lo menos dos propiedades, cada una medida por lo menos por dos instrumentos.

En vistas de las distinciones entre los tres tipos de validez, ¿debería usarse cada uno de los tests cuando se evalúa la validez de un determinado instrumento de medición? No hay una solución simple a este problema. Su significado lleva a un equipo de expertos de diferentes disciplinas a recomendar que el examen completo de un instrumento de medición incluye información sobre los tres tipos de validez.¹⁰ Así

⁸ Lee J. Cronbach y Paul Meehl, "Construct Validity in Psychological Tests," *Psychological Bulletin*, 52 (1955): 281-302.

⁹ Donald Campbell y Donald W. Fiske, "Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix," *Psychological Bulletin*, 56 (1959): 81-105.

¹⁰ Véase American Psychological Association Committee on Psychological Tests, "Technical Recommendations for Psychological Tests and Diagnostic Techniques," *Psychological Bulletin Suppl.*, 51

en la primera fase de la construcción de una medida, uno debe evaluar teorías que sirvan como fundamento al instrumento; luego debe definirse una población satisfactoria de ítems, los cuales deben ilustrar un muestreo representativo; y finalmente, la validez predictiva del instrumento debe ser juzgada correlacionándolo con un criterio externo.

Confiabilidad

La confiabilidad es una preocupación central para los científicos sociales porque los instrumentos de medición raramente son completamente válidos. En muchos casos, la evidencia de validez es casi nula; uno tiene que evaluar el instrumento de medición respecto a otras características y asumir su validez. Un método frecuentemente usado para evaluar un instrumento es su grado de confiabilidad.

La **confiabilidad** refiere al punto hasta el cual el instrumento de medición contiene *errores variables*, esto es, errores que difieren de una observación a otra durante cualquier instancia de medición o que varían de tiempo en tiempo para una determinada unidad de análisis medida dos o más veces por el mismo instrumento. Por ejemplo, si uno mide el largo de un escritorio en dos puntos temporales con el mismo instrumento –digamos, una regla- y obtiene resultados apenas diferentes, el instrumento contiene errores variables. Debido a la naturaleza indirecta de la mediciones en las ciencias sociales, los errores que ocurren cuando se miden variables sociales tienden a ser mayores a los que ocurren cuando se miden variables físicas. Factores tales como la distracción momentánea al completar un cuestionario, o las instrucciones ambiguas, o las dificultades técnicas (un lápiz se rompe mientras el entrevistado llena un formulario) pueden causar la introducción de errores variables (de medición). Estos errores se llaman errores variables porque la cantidad de errores varía de una observación a la siguiente y también porque la cantidad de errores para una determinada observación es diferente cada vez que es medida.

Cada medición, entonces, contiene dos componentes: un *componente de verdad* y un *componente de error*. La confiabilidad se define como la razón de varianza del resultado de verdad sobre la varianza total de los resultados medidos.¹¹ (La varianza es una medida de la variación de las observaciones, o resultados; es una descripción del punto hasta el cual las observaciones difieren mutuamente; esto es,

$$\sigma^2 = \frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N}$$

por información detallada véase el capítulo 15.) Algebraicamente, el resultado observado de cada persona se puede representar como

$$x_i = t_i + e_i \quad (7.1)$$

donde x_i = el resultado obtenido por la persona i

(1954) pt. 2: 1-38, and Donald T. Campbell, Recommendations for APA Test Standards Regarding Construct, Trait, or Discriminant Validity," *American Psychologist*, 15 (1960): 546-553.

¹¹ Esta definición y la siguiente presentación están basadas en C. G. Helmstadter, *Research Concepts*, pp. 169-176.

t_i = el resultado real de la persona i
 e_i = la cantidad de error que le ocurre a la persona i en el momento en que se lleva a cabo la medición

Expresado en términos de varianza, tenemos

$$\sigma_x^2 = \sigma_t^2 + \sigma_e^2$$

donde σ_x^2 = la varianza de los resultados observados
 σ_t^2 = la varianza de los resultados reales
 σ_e^2 = la varianza de los errores

La confiabilidad, definida como la razón de varianza de los resultados reales sobre la varianza de los resultados observados, puede ser expresada como

$$\frac{\sigma_t^2}{\sigma_x^2} = \frac{\sigma_x^2 - \sigma_e^2}{\sigma_x^2} \quad \text{Confiabilidad} = \quad = \quad (7.2)$$

En la Ecuación (7.2) podemos ver que si la medición no contiene nada más que errores, entonces $\sigma_x^2 = \sigma_e^2$ y la confiabilidad es de cero. De todas formas, cuando no hay ningún error variable, $\sigma_e^2 = 0$, y la razón definida como confiabilidad resulta

$$\frac{\sigma_x^2}{\sigma_x^2} = 1$$

La medida de confiabilidad varía en una escala de 0 a 1, teniendo el primer valor cuando la medición contiene sólo errores y alcanzando 1 cuando no hay ningún error variable en la medición.

En la práctica, es imposible computar el resultado real independientemente de la cantidad de error que ocurre en cualquier medición particular. En consecuencia, la razón σ_t^2 / σ_x^2 tiene que ser estimada. Hay tres modos comunes de estimación de la confiabilidad: el método de test-retest, la técnica de las formas paralelas, y el método de dividir a la mitad.

Método de test-retest

El **método de test-retest** deriva directamente de la definición conceptual de la confiabilidad. Un instrumento de medición es administrado al mismo grupo de personas en dos ocasiones diferentes, y la correlación entre los dos conjuntos de observaciones (resultados) es computada. El coeficiente obtenido es la *confiabilidad estimada*. Con este método, el error es definido como cualquier cosa que lleve a una persona a obtener un resultado diferente de una medición a otra. Simbólicamente,

$$r_{xx'} = \frac{S_t^2}{S_x^2} \quad (7.3)$$

donde x = el rendimiento en la primer medición
 x' = el rendimiento en la segunda medición
 $r_{xx'}$ = el coeficiente de correlación entre x y x'
 S_t^2 = la varianza estimada de los resultados reales
 S_x^2 = la varianza calculada de los resultados observados

La correlación $r_{xx'}$ provee una estimación de confiabilidad definida como la razón de la varianza real sobre la varianza observada. (Por métodos de computación del coeficiente de correlación, véase el capítulo 16.)

El método de test-retest tiene dos limitaciones centrales. Primero, la medición en una ocasión puede influenciar las mediciones en ocasiones subsecuentes. Si, por ejemplo, el instrumento es un cuestionario, un entrevistado puede recordar preguntas específicas y las respuestas en el mismo sentido que en la primera ocasión, y así producir una estimación de confiabilidad alta pero exagerada. En segundo lugar, muchos fenómenos cambian a través del tiempo. Es posible que los cambios ocurran en la variable medida durante el intervalo de las mediciones, disminuyendo así la estimación de confiabilidad. El método de test-retest, entonces, puede tanto exagerar como subestimar la confiabilidad real del instrumento, y en muchos casos es difícil determinar qué ha ocurrido.

Técnica de las formas paralelas

Una forma de contar con las dos limitaciones del método de test-retest es usando la **técnica de las formas paralelas**. Esta técnica requiere de dos formas paralelas del instrumento de medición. Las dos formas entonces son administradas a un grupo de personas, y los dos conjuntos de medidas (resultados) son correlacionados para obtener una estimación de confiabilidad. Con esta técnica, existe el problema de determinar si las dos formas del instrumento en verdad son paralelas. Aunque los test estadísticos han sido desarrollados para determinar si las formas son paralelas en términos de mediciones estadísticas, la evaluación de los contenidos de las dos formas es subjetiva.¹²

Método de dividir a la mitad

El **método de dividir a la mitad** estima la confiabilidad mediante el tratamiento de cada una de las dos o más partes del instrumento de medición como una escala separada. Supongamos que el instrumento de medición es un cuestionario. El cuestionario se divide en dos grupos, usando las preguntas impares para un grupo y las preguntas pares para el otro. Cada uno de los dos conjuntos de preguntas es tratado separadamente y medido correspondientemente. Los dos conjuntos son entonces correlacionados, y esto se toma como una estimación de confiabilidad. Para corregir el coeficiente de correlación obtenido entre las dos mitades, se aplica la siguiente fórmula, conocida como la fórmula de la profecía de Spearman-Brown:

$$r_{xx'} = \frac{2r_{oe}}{1 + r_{oe}} \quad (7.4)$$

donde $r_{xx'}$ = la confiabilidad del test original

¹² Véase Harold Gulliksen, *Theory of Mental Tests* (New Cork: Wiley, 1962).

r_{oe} = el coeficiente de confiabilidad obtenido mediante la correlación entre los resultados de las proposiciones impares y los resultados de las proposiciones pares

Esta correlación asume que un instrumento que contiene una cantidad de preguntas $2n$ será más confiable que un instrumento que contiene una cantidad de preguntas n y que porque el largo del instrumento ha sido reducido dividiéndolo entre impares y pares, el total del instrumento tendrá mayor confiabilidad que cualquiera de sus mitades.

Cronbach, Rajaratnam, y Glesser introdujeron una revisión al concepto tradicional de confiabilidad.¹³ Estos autores sostenían que la preocupación central de la teoría de la confiabilidad es la de responder a la pregunta “¿A qué universo de potenciales mediciones nos gustaría generalizar? Así en lugar de confiabilidad, se introduce la idea de generalización. La **generalización** implica que lo que uno en realidad quiere conocer acerca de un conjunto de mediciones es ¿hasta qué punto y con respecto a qué propiedades es igual a otros conjuntos de mediciones que uno pudo haber utilizado de un determinado universo de mediciones potenciales? ¿Y hasta qué punto y con respecto a qué propiedades es diferente de otras mediciones que uno pudo haber ilustrado de aquel universo de mediciones potenciales? Si uno formula las preguntas sobre mediciones potenciales similares y diferentes, uno pregunta por los límites de generalización de los resultados de su propio conjunto de mediciones. Si consideramos una relación particular entre mediciones como evidencia de la confiabilidad o generalidad depende de cómo elijamos definir la similitud y la diferencia de condiciones y mediciones. La construcción de lo que es lo mismo y de lo que es diferente en conjuntos de medición depende, siempre, del problema de investigación.¹⁴

Resumen

1. La medición es la asignación de numeraciones a las variables, propiedades, o eventos de acuerdo a reglas. El concepto más significativo en esta definición es “reglas”. La función de una regla es atar el procedimiento de medición a la realidad; establecer el isomorfismo entre una cierta estructura numérica y la estructura de las variables medidas. Al establecer el isomorfismo, los investigadores pueden desarrollar un análisis cuantitativo con las numeraciones que representan a las propiedades. El isomorfismo entre los sistemas numéricos y las propiedades empíricas lleva a una distinción entre cuatro niveles de medición: nominal, ordinal, intervalar, y razón. En general, el análisis cuantitativo permitido para un determinado conjunto de números depende del nivel de medición obtenido.

2. Las ideas de validez y confiabilidad son inseparables de la teoría de la medición. La validez se ocupa de la pregunta acerca de si uno está midiendo lo que uno piensa que está midiendo. Tradicionalmente, se distinguen tres tipos básicos de validez, cada uno de los cuales se relaciona con un aspecto diferente de la situación de medición: validez de contenido, validez empírica, y validez de constructo. Para validar un determinado instrumento de medición, uno debe buscar información acerca de estos tres tipos. La confiabilidad indica el punto hasta el cual una medición contiene errores variables. Operacionalmente, se asume que cualquier medida contiene un

¹³ Lee J. Cronbach, Nageswars Rajaratnam, y Goldine C. Glesser, “A Theory of Generalizability: A Liberalization of Reliability Theory,” *British Journal of Statistical Psychology*, 16 (1963): 137-163.

¹⁴ Por la expresión estadística de índice de generalidad, véase *ibid.*, y Goldine C. Glesser, Lee J. Cronbach, y Nageswars Rajaratnam “Generalizability of Scores Influenced by Multiple Scores of Variante,” *Psychometrika*, 30 (1965): 395-418.

componente de verdad y un componente de error y que la proporción del monto de variación en el componente de verdad sobre la variación total mide la confiabilidad. Esta medida puede ser estimada mediante uno o más de los siguientes métodos: test-retest, formas-paralelas, y división-en-mitades. La noción de generalidad implica que la preocupación central de la confiabilidad tiene que ver con hasta qué punto un conjunto de mediciones es similar a otros conjuntos de mediciones que podrían haber sido ilustrados de un determinado universo potencial de mediciones.

Términos clave para la revisión

Medición	Validez de muestreo
Isomorfismo	Validez predictiva
Indicador	Validez de constructo
Nivel nominal	Técnica de grupos-conocidos
Nivel ordinal	Confiabilidad
Nivel intervalar	Medida de confiabilidad
Nivel razón	Método de test-retest
Errores de medición	Técnica de formas-paralelas
Validez	Método de división-en-mitades
Validez aparente	Generalización

Preguntas de estudio

1. Defina la medición y explique por qué la medición es importante para la investigación científica.
2. ¿Cuáles son los niveles de medición? ¿Por qué son importantes las diferencias entre los niveles de medición?
3. Defina el concepto “validez” y explique cómo distinguir entre los tres mayores tipos de validez.
4. Defina el concepto “confiabilidad” y discuta los modos de alcanzarla.

