



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

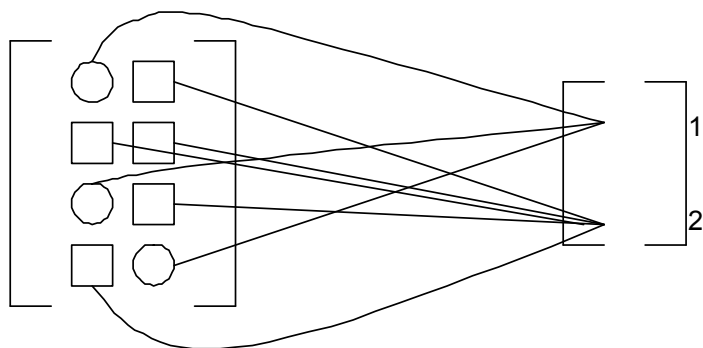
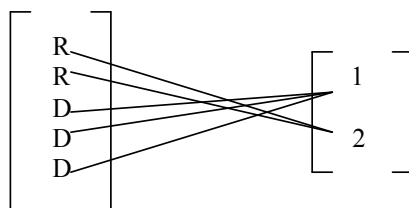


Figura 7.2
Asignación según la aplicación de una regla



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.

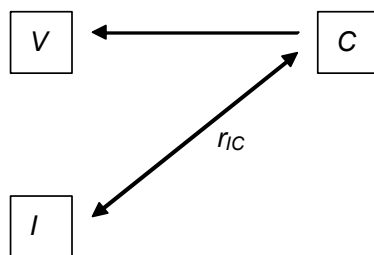
Validez empírica

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.

CAPÍTULO 9

Métodos Observacionales

En este capítulo, primero, discutimos y ejemplificamos la idea de triangulación. Luego, discutimos los varios papeles que tiene la observación en la investigación en ciencias sociales. Finalmente, presentamos las estrategias para conducir observaciones directas, seguido de una discusión sobre la observación controlada en el laboratorio y en el campo.

Una vez que se ha decidido el “qué” y el “cómo” de una investigación se procede a la recolección de los datos. Los datos de las ciencias sociales se obtienen cuando los investigadores registran observaciones sobre el fenómeno a estudiar. Existen cuatro formas de recolección de datos: métodos de observación (observational methods), investigación por encuesta (survey research), análisis de datos secundarios e investigación cualitativa. Cada forma emplea ciertos métodos específicos. En los próximos capítulos se discuten los más importantes. Cabe aclarar que cada una de estas formas presenta ciertas ventajas únicas, pero también ciertas limitaciones que les son inherentes. Por este motivo, resulta muy útil -cuando es factible- la triangulación de métodos, es decir, usar más de una forma de recolección de datos para testear la misma hipótesis.

Triangulación

En ciencias sociales, los datos se obtienen tanto en escenarios formales como informales e involucran tanto actos verbales (orales y escritos) como no verbales. Las principales cuatro formas de recolección de datos se derivan de la combinación de esos dos escenarios y esos dos tipos de actos: métodos de observación (observational methods), investigación por encuesta (survey research), análisis de datos secundarios e investigación cualitativa. En un extremo, donde la preocupación son las acciones no verbales en escenarios informales, la observación participante -una forma de investigación cualitativa- es un método común de recolección de datos. En el otro extremo, donde la investigación se enfoca en los actos verbales (orales y escritos) en escenarios formales y estructurados, las formas más comunes de recolección de datos son los experimentos de laboratorio y los cuestionarios estructurados.

Como aclaramos, cada uno de estos métodos de recolección tiene ventajas y limitaciones. Por ejemplo, si observamos comportamientos mientras ocurren (observación directa), podemos no conocer las razones por las que ocurren (lo cual puede conocerse a través de cuestionarios estructurados). Asimismo, si preguntamos a los entrevistados que den cuenta de sus comportamientos verbalmente (entrevistas) no tenemos garantías de que su comportamiento real (estudiado a través de observaciones directas o registros existentes) sea idéntico al comportamiento que “dicen” haber tenido. Por ejemplo, en un estudio sobre la validez de las respuestas a las preguntas sobre el voto de la encuesta a las madres asistidas, Weiss afirmó:

Acerca de las preguntas sobre el voto y la inscripción, un 82% de las madres asistidas respondió acertadamente. Un 16% reportó su inscripción de más (overreported) y 2% reportó su inscripción de menos. La cantidad y la dirección del error de las respuestas son similares a aquellas de las amplias poblaciones de clase media cuyos auto reportes sobre el voto habían sido previamente validados en estudios previos.

Tanto en el comportamiento del voto como en otro tipo de comportamientos, hay un discrepancia entre lo que la gente “dice” y su comportamiento real.

Hasta cierto punto, los descubrimientos de la investigación se ven afectados por la naturaleza del método de recolección de datos usado. Los descubrimientos que están fuertemente afectados por el método usado podrían ser artefactos en vez de datos empíricos.

Para reducir el grado de especificidad en los cuerpos de conocimiento, un investigador puede usar dos o más métodos de recolección de datos par testear hipótesis y medir variables; esta es la esencia de la **triangulación**. Por ejemplo, un cuestionario estructurado puede ser complementado con una entrevista en profundidad, registros existentes o observación de campo. Los descubrimientos coincidentes entre distintos métodos aumenta la credibilidad de los descubrimientos de investigación. Y como estrategia de investigación, la triangulación eleva a los científicos sociales “por encima de las bases personales que provienen de la metodologías individuales. Al combinar métodos e investigadores en el mismo estudio, los observadores pueden superar parcialmente las deficiencias que provienen de un investigador o de un método”.

Roles de Observación

La investigación social está apoyada en la observación. Los científicos políticos observan, entre otras cosas, el comportamiento de quienes ocupan roles políticos; los antropólogos observan rituales en sociedades simples; y los psicólogos sociales observan interacciones en pequeños grupos. En un sentido, toda la ciencia social comienza y termina con observaciones.

El mayor beneficio de la observación es que hace posible estudiar el comportamiento mientras ocurre. El investigador no tiene que preguntar a la gente sobre su propio comportamiento y las acciones de otros, el o ella simplemente pueden mirarlos hacer y decir cosas. Esto, a cambio, permite la recolección de datos de primera mano que no están contaminados por factores que están entre el investigador y el objeto de estudio. Por ejemplo, cuando a la gente se le pide que reporte su comportamiento pasado, distorsiones en la memoria pueden contaminar los datos, en cambio la memoria no tiene efecto sobre los datos del comportamiento recolectados a través de métodos observacionales.

Además, los datos recolectados a través de la observación pueden describir los fenómenos observados mientras ocurren en sus escenarios naturales. Otro método de recolección de datos introduce elementos de artificialidad en el ambiente de la investigación. Una entrevista, por ejemplo, es una forma de interacción cara a cara, sujeta a problemas peculiares por la falta de consenso alrededor de los roles del investigador y el entrevistado. En dicha interacción, los entrevistados pueden actuar de modos que no son comunes a su comportamiento típico. La artificialidad puede reducirse en los estudios basados en la observación, especialmente cuando los sujetos no son conscientes de ser observados o cuando se acostumbran al observador y no lo consideran un intruso.

Algunos estudios se enfocan en individuos que no son capaces de dar respuestas verbales o articular lo que quieren decir de manera que tenga sentido. Por ejemplo, es necesario usar la observación en estudios sobre niños porque es difícil que los niños puedan ser introspectivos, verbalizar, y permanecer atentos durante largas tareas.

Los métodos de observación también pueden ser usados para personas que no desean expresarse verbalmente. La observación demanda menos compromiso por parte del individuo que está siendo estudiado que los reportes verbales. A su vez, los

reportes verbales pueden ser validados y comparados con el comportamiento a través de observaciones. Finalmente, la relación entre una persona y su ambiente no se ve alterada en los estudios de observación. Las oportunidades para analizar el *contexto* del comportamiento mejoran gracias a la habilidad del investigador para observar el impacto del ambiente en los individuos investigados.

La observación toma diversas formas. Ésta incluye experiencias casuales así como artículos de laboratorio como pantallas y cámaras de video. Las distintas formas de observación hacen que sea un método apropiado para una variedad de objetivos de investigación. Puede usarse en investigación exploratoria para obtener indicios que luego pueden ser testeados como hipótesis. Los métodos observacionales también pueden usarse para recolectar datos suplementarios que pueden ayudar a interpretar o calificar los descubrimientos obtenidos por otros métodos, o pueden ser usados como métodos primarios de recolección de datos en estudios descriptivos.

La observación puede tener lugar en escenarios naturales o en el laboratorio. Un problema de investigación como los patrones de aprendizaje puede ser estudiado mientras ocurre en una situación de vida "real" (por ejemplo, una clase en un lugar de juegos) o en el cuarto de laboratorio. Los procedimientos observacionales pueden variar entre una flexibilidad total guiados por un problema general hasta el uso de instrumentos estructurados diseñados con anterioridad. Los investigadores pueden participar en las actividades del grupo que están observando; pueden ser vistos como miembros del grupo pero minimizar su participación en él; deben asumir roles de observadores sin hacer parte del grupo; o su presencia puede ser ocultada de las personas a las que observa. Cualquiera sea el propósito del estudio y el procedimiento observacional usado, hay que tener en cuenta tres consideraciones principales si se pretende que los datos obtenidos sean sistemáticos y significativos: qué observar, cuando observarlo y cómo registrarlos, y cuánta inferencia se requiere.

Tipos de comportamiento

La primera consideración importante se refiere a *lo que debe ser observado*. Suponiendo que una persona interesada en estudiar la relación entre frustración y agresión realiza una hipótesis en donde la frustración conlleva a la agresión, para corroborar esta hipótesis es necesario observar la frustración y la agresión. Esto requiere una definición operacional clara y precisa de las dos variables. La medida de las variables "frustración" y "agresión", y de otras variables, se puede basar en un comportamiento no verbal espacial extra lingüístico o lingüístico.

Comportamiento No Verbal

El comportamiento no verbal es el movimiento corporal del organismo y consiste en expresiones motoras que se pueden originar en varias partes del cuerpo. El comportamiento no verbal ha sido estudiado extensivamente, y ha sido mostrado repetidamente como un indicador válido de procesos sociales, políticos y psicológicos. Paul Ekman sugiere que las observaciones del comportamiento no verbal generan datos que pueden servir "para repetir, contradecir o sustituir un mensaje verbal, como también para acentuar ciertas palabras, manteniendo el flujo comunicativo, reflejar cambios en una relación en asociación con mensajes verbales particulares e indicar el sentimiento de una persona acerca de su afirmación verbal" (Paul Ekman; 1957). Expresiones faciales, en particular, expresan un rango de emociones, incluyendo miedo, sorpresa, enojo, disgusto y soledad.

El comportamiento espacial se refiere a los intentos de los individuos para estructurar el espacio que los rodea. Por ejemplo, la gente se mueve hacia delante, se

mueve desde, mantiene cercanía y mantiene distancia. El rango, la frecuencia y resultados de estos movimientos proveen datos significativos para una variedad de propósitos de investigación.

Por ejemplo, hay distintos patrones en la forma que la gente usa el espacio que la rodea cuando interactúa con otros. En diferentes culturas hay costumbres no escritas regulando cuanta distancia los individuos dejan entre sí cuando se acercan: los latinoamericanos tienen espacios personales más cercanos que los norteamericanos, alemanes e ingleses. Las variaciones culturales en el espacio personal tienen serias implicaciones en sociedades o ciudades culturalmente heterogéneas. Un individuo proveniente de un contexto alemán puede percibir como incómoda la interacción con un latinoamericano porque tienen dificultad para establecer espacios interpersonales satisfactorios. Se puede generar fricción entre dos personas ya que alguno puede ver al otro como grosero. De hecho, ambos están atentando contra el espacio establecido como cómodo y aceptable para el otro en su respectiva cultura. Hasta dentro de una sociedad hay diferencias en las normas de espacio personales. Haillo y Thomson observan que en la niñez, los negros interactúan de manera más cercana que los blancos. Sin embargo, en la adolescencia, los negros interactúan a una distancia mayor que los blancos (Haillo y Thomson, 1980). Una respuesta común a las violaciones del espacio personal es el stress. Las personas cuyo espacio personal es violado reportan sentirse tensas y ansiosas. Las funciones psicológicas como la respuestas en la piel, el ritmo cardíaco y la presión arterial incrementan cuando un espacio personal es violado.

Comportamiento extralingüístico

Las palabras o el contenido lingüístico constituyen sólo una pequeña porción del comportamiento. El comportamiento sin contenido lingüístico (“noncontent behaviors”) como el ritmo y el volumen de la conversación, la tendencia a interrumpir y la pronunciación constituyen una rica fuente de datos y son generalmente conocidos como comportamiento extra lingüístico o paralenguaje. El significado del paralenguaje en el estudio del comportamiento humano ha sido documentado en numerosos estudios. Por ejemplo, una característica vocal como el tono mide exactamente estados emocionales. La medida promedio del discurso espontáneo incrementa a medida que el tamaño del grupo incrementa (Soskin y Vera, 1963). La frecuencia de las interrupciones refleja diferencias en el poder personal. Emociones pasivas como la tristeza son expresadas a través del discurso lento y de bajo volumen y tono, y las emociones activas como el enojo son comunicadas por discursos rápidos, de volumen y tono altos. Estos ejemplos solo dan una pista del rango de aplicaciones de indicadores extra lingüístico en el estudio del comportamiento y demuestran el significado potencial del comportamiento sin contenido para la investigación social.

Comportamiento Lingüístico

El comportamiento lingüístico es el contenido manifiesto del discurso y las características estructurales del habla. La medición del comportamiento lingüístico ha sido usada ampliamente en estudios sobre la interacción social. Robert Vales, por ejemplo, desarrolló un conocido sistema para organizar y codificar el proceso de interacción en grupos involucrados en actividades de resolución de problemas. El sistema de Vales, análisis de procesos de interacción (IPA), contiene 12 tipos de comportamientos distintivos dentro de los cuales la interacción de los miembros de un grupo puede ser codificada y analizada. El código de categorías IPA es mostrado en el Cuadro 9.1.

Tiempo y Registro

La segunda consideración importante en los estudios de observación se refiere al tiempo y el registro de las observaciones. Obviamente, es imposible hacer un infinito número de observaciones, entonces es necesario tomar una decisión sobre cuando observar. Un acercamiento aceptable a este problema es seguir una muestra de distintos puntos en el tiempo ("time sampling schedule"). "Time sampling" se refiere a la selección de unidades de observación en distintos puntos del tiempo. Las unidades de observación pueden ser elegidas de modos sistemáticos para asegurar la representación de una cierta cantidad de comportamientos. Por ejemplo, uno puede hacer una observación durante un período de 15 minutos de cada hora seleccionada al azar después de estratificar el día de la semana y la hora del día. Otro procedimiento útil de muestreo ("sampling") es el muestreo de individuos (*individual sampling*), también llamado *specimen records*. El investigador selecciona un individuo y registra todo el comportamiento y los eventos centrados en ese individuo. Por ejemplo, el observador selecciona un chico y registra todos los ejemplos de agresión física entre el chico y sus compañeros. Cada 30 minutos se selecciona un chico diferente. Los datos recolectados luego representan el comportamiento de todos los chicos en la clase.

Además de desarrollar un diseño del "time sampling", el investigador tiene que desarrollar un sistema de códigos para registrar las observaciones. La categorización es necesaria para transformar la complejidad de los eventos en datos que puedan ser expresados numéricamente o cuantificados. Este sistema de códigos puede ser construido tanto a través de un enfoque deductivo como inductivo. Un enfoque deductivo implica que el investigador comience con una definición conceptual, luego especifique los indicadores del comportamiento que debe ser observado y luego estandarice y valide los instrumentos resultantes. El enfoque deductivo es implementado cuando las observaciones son asignadas a categorías mientras se hace el registro. Opuestamente, el enfoque inductivo requiere primero de la selección de indicadores y pospone las definiciones hasta que se identifique cualquier patrón. Cada enfoque involucra algún tipo de riesgo. Con el enfoque deductivo, es difícil prever si la definición conceptual es correcta. El enfoque empírico presenta dificultades para interpretar las observaciones. La mejor forma para reducir este riesgo es combinar los dos enfoques. K. Weicks sugiere que

en la secuencia ideal, el observador debería comenzar con el enfoque empírico, obtener un amplio registro de los eventos naturales, inducir algunos conceptos de los registros y luego recolectar un segundo grupo de registros que sean más específicos y se dirijan más directamente a los conceptos inductivos. (Weick)

Más allá de que se utilice un método deductivo o inductivo, las categorías a las cuales son asignadas las observaciones deben exhibir ciertas características. Un sistema de categorías debe

limitar la observación a un segmento o aspecto de... comportamiento, y construir un conjunto de categorías finito en el cual una y solo una de cada unidad observada puede clasificarse. El registro obtenido pretende mostrar, para cada período de observación, el número total de unidades de comportamiento que ocurren y el número clasificable en cada categoría (Donald M., Mitzel, H. E.; 1963)

En otras palabras, las categorías deben ser explícitas, exhaustivas, y mutuamente excluyentes. Una categoría explícita es especificada en términos del acontecimiento a ser observado, la situación en la cual el acontecimiento tiene lugar y el evento que precede o sigue a los acontecimientos observados.

Inferencia

El tercer tipo de consideración en los estudios sobre estructura observacional se relaciona con el grado de inferencia hecho por el observador. La mayoría de los registros en observación involucran inferencias. Un investigador observa un cierto acto o comportamiento y debe procesar esta observación e inferir que el comportamiento mide una cierta variable. Algunos sistemas observacionales requieren un bajo grado de inferencia observacional, por ejemplo, actos tales como “formulé una pregunta”, “sugerí un curso de acción”, e “interrumpí a otro miembro del grupo”. Muchos actos, sin embargo, requieren un grado de inferencia. Supongamos que uno observa a un adulto pegándole a un niño. Debe hacerse una inferencia sobre si este acto representa “agresión”, “comportamiento agresivo”, “hostilidad”, “violencia” u otra variable. La exactitud de dicha inferencia depende en gran medida de la competencia del observador. En las mismas condiciones, los observadores bien entrenados tienden a realizar inferencias más confiables que los menos entrenados.

Como medio para incrementar la confianza de las inferencias, se diseñaron programas de entrenamiento aplicables a varias situaciones observacionales. Generalmente, un programa comienza con la exposición de una teoría, la investigación de ciertas hipótesis involucradas en un estudio dado y la explicación del sistema de categorías construido para registrar las observaciones. Después de que los aprendices tuvieron la oportunidad de elevar preguntas, aplican el sistema de categorías en una situación real de vida, y sólo entonces comienza la recolección de datos.

Tipos de Observación

La medida en que las decisiones referidas al comportamiento, tiempo, registro e inferencia son sistemáticamente y rigurosamente implementadas es un criterio por el cual podemos distinguir entre sistemas observacionales controlados y no controlados. Un sistema observacional controlado es caracterizado por decisiones claras y explícitas sobre qué, cómo y donde observar; un sistema no controlado es considerablemente menos sistemático y permite mayor flexibilidad. Por ejemplo, en una **observación controlada**, se esboza un “time sample” antes de la observación; en la **observación no controlada**, casi nunca se toman muestras (“samples”). La elección entre observación controlada y no controlada depende del problema de investigación y el diseño de la investigación; es decir, la observación controlada es frecuentemente usada con diseños de investigación experimental y casi nunca con diseños pre experimentales o estudios cualitativos. En este capítulo se discute sobre los sistemas de observación controlada, en el capítulo 12 se presentan las observaciones no controladas (investigación cualitativa).

Observaciones Controladas

Las observaciones controladas son llevadas a cabo tanto en el laboratorio como en el campo. En ambos escenarios, el investigador desea inferir la causalidad maximizando el control sobre las variables extrínsecas e intrínsecas mientras emplea uno de los varios diseños de investigación experimental y registra las observaciones sistemáticamente.

Experimentaciones de Laboratorio

El método de recolección de datos más controlado en ciencias sociales es el laboratorio de experimentación. Éste involucra la creación de condiciones en un ambiente controlado (el laboratorio) que simula ciertas características de un ambiente natural, y una manipulación supervisada de una o más variables independientes por vez para observar los efectos producidos.

Ejemplos clásicos de experimentación de laboratorio son los experimentos de Solomon Asch sobre influencia interpersonal. El objetivo de Asch era examinar las condiciones sociales y personales que inducen a los individuos a aceptar o resistir las presiones del grupo cuando dichas presiones son percibidas como contrarias al hecho. Asch desarrolló un procedimiento para colocar a los individuos en fuerte desacuerdo con sus pares para medir el efecto de esta relación en ellos. Ocho individuos fueron instruidos para marcar la longitud de una cierta línea con algunas pulgadas de diferencia. Se pidió a los miembros del grupo que anuncien su juicio en voz alta. En el medio del test, un individuo de pronto se contradijo con el resto del grupo, porque Asch instruyó a los otros siete miembros para que respondan en ciertos puntos erróneamente. Los errores de la mayoría fueron amplios, en un rango de entre $\frac{1}{2}$ y $\frac{3}{4}$ pulgadas. Los ocho individuos confrontaron una situación en la cual el grupo unánimemente contradijo la evidencia de sus sentidos. Este individuo, comúnmente llamado “*el sujeto crítico*” fue el objeto de la investigación. Asch también utilizó un grupo de control donde los errores introducidos por la mayoría no fueron del mismo orden encontrados bajo las condiciones experimentales. Uno de los descubrimientos interesantes fue un marcado movimiento hacia la mayoría.

El experimento de Asch ejemplifica las dos mayores ventajas de la experimentación en laboratorio: permite un riguroso control sobre factores extrínsecos e intrínsecos, y provee evidencia clara sobre la causalidad. Asch eliminó los efectos de muchas variables que podían haber causado que los sujetos críticos produzcan o se opongan a la presión del grupo; esto aumentó la posibilidad de observar diferencias existentes debido a su tratamiento experimental. Además, Asch pudo especificar sin ambigüedades qué causó el movimiento de sus sujetos críticos hacia la mayoría porque él mismo controló y manipuló la variable independiente –los siete miembros del grupo a los que se les dijo que respondan equivocadamente. A su vez, Asch varió el tratamiento sistemático en un modo sistemático, permitiendo así la especificación precisa de las diferencias importantes. Finalmente, el experimento se construyó de forma que permitió una clara detección de los efectos del tratamiento experimental: los sujetos críticos tenían que emitir su juicio en voz alta. Tenían que declarar y tomar una posición definitiva vis a vis con sus pares. No podían evitar el dilema indicando condiciones externas a la situación experimental.

Los experimentos de laboratorio varían en complejidad y diseño, dependiendo de los problemas de investigación y de la ingenuidad del experimentador. Los experimentadores tienen que construir un conjunto de procedimientos que capturen el significado de su conceptualización y que les permita testear las hipótesis. Esto, a su vez, demanda la invención de un método que mida el efecto que esto tiene en el comportamiento de los individuos analizados y en la construcción de un escenario dentro del cual las manipulaciones básicas de las variables independientes tiene sentido y sus mediciones son válidas y confiables.

Realismo Experimental y Mundano

Es posible cuestionar el sentido de la experimentación de laboratorio, dado que no representa una situación del mundo real. En el experimento de Asch, los sujetos críticos juzgaron un evento físico claro (la longitud de las líneas) y fueron

contradecidos por sus pares. Sin embargo, en la vida cotidiana, es poco probable una situación donde la evidencia clara del propio sentido es contradicha por los juicios unánimes de los pares.

Este problema condujo a una distinción entre dos sentidos en los cuales cualquier experimento puede considerarse realista. En un sentido, un experimento es realístico si la situación es realística para los participantes en la investigación, si los involucra y los afecta. Este tipo de realismo se denomina comúnmente “realismo experimental”. En el experimento de Asch, los sujetos críticos exhibían signos de tensión y ansiedad. Los sujetos estaban actuando a una situación que era tan real para ellos como cualquier otra experiencia fuera del laboratorio.

El segundo sentido del realismo, se refiere a la medida en la cual los eventos que ocurren en un escenario de laboratorio tienen probabilidad de ocurrir en el mundo real. Este tipo de realismo se llama “realismo mundano”. Un experimento que es alto en realismo mundano y bajo en realismo experimental no necesariamente arroja resultados más significativos que un experimento que es alto en realismo experimental y bajo en realismo mundano. El realismo experimental habilita al experimentador a aumentar la validez interna del experimento al producir un efecto significativo dentro de la situación experimental.

Fuentes de sesgo en Experimentos de Laboratorio

A pesar de las ventajas de los experimentos de laboratorio, ellos tienen ciertas limitaciones inherentes. Estas pueden clasificarse en tres tipos: sesgos (“bias”) debido a las características de la demanda de la situación experimental en sí misma, sesgos debido a la influencia no intencionada de los experimentadores, y los artefactos de medida.

Características de demanda. Los sesgos debidos a características de demanda pueden ocurrir cuando los individuos saben que están en una situación experimental, son conscientes de que están siendo observados y creen que se espera de ellos ciertas respuestas. Consecuentemente, pueden no responder a la manipulación de la experimentación tal y como se espera, sino a su interpretación de las respuestas que se supone que estas manipulaciones tienen que provocar.

Aunque el experimentador anuncie que no hay respuestas correctas o incorrectas los sujetos igualmente asumirán que se espera de ellos ciertos comportamientos y tratarán de presentar una imagen lo más positiva posible. Los sujetos pueden descubrir la hipótesis de investigación y responder de un modo consistente con ella en un intento por cumplir con el deseo del experimentador. Una práctica comúnmente usada para contrarrestar este tipo de sesgo es reducir la conciencia de los participantes de que están siendo observados en una situación de investigación. Otra estrategia es discutir sólo lo general en vez de los objetivos específicos de investigación. Lo que se cree es que si varios sujetos modifican su comportamiento de modo de apoyar o refutar una hipótesis errónea, los resultados referidos a las hipótesis reales pueden no verse afectados, al menos no sistemáticamente.

Sesgos del experimentador El comportamiento del lado del experimentador que se supone que no es parte de la manipulación experimental, pero sin embargo, afecta a los participantes es denominado sesgo experimental o “efecto de la expectativa experimental”. Los experimentadores que son conscientes de los efectos que desean de los individuos pueden comunicar sus expectativas involuntariamente de varias formas, por ejemplo, mostrando tensión o alivio en ciertas ocasiones o afirmando con la cabeza. El sesgo del experimentador es el resultado de las motivaciones del observador. Usar grabadores, cámaras de otro tipo de procesos para

minimizar las interacciones entre el experimentador y los participantes puede minimizar de modo no intencional los sesgos del experimentador y eliminar la comunicación de las expectativas. Los efectos de “bias” también fueron mitigados al usar experimentadores con expectativas diferenciales sobre el resultado final de la investigación. En un estudio se incluyeron experimentos con distintas expectativas sobre los efectos de las variables manipuladas como una de las variables en el diseño experimental. En este caso, los investigadores analizaron hasta qué punto sus diferentes expectativas produjeron diferentes resultados. También se recomienda el uso de más de un observador o recolector de datos para estimar los efectos de los rasgos de personalidad, las características físicas y las diferencias sutiles en el tratamiento de los participantes.

Artefactos de medida. La medición es una parte crucial del proceso de investigación. En los experimentos de laboratorio donde los efectos de una variable independiente pueden ser pequeños, cortos y sensibles, se requiere de mediciones precisas para detectar dichos efectos. A su vez, los procedimientos de medición no son independientes de los problemas de diseño de investigación, dado que los procesos de medición pueden crear interpretaciones adicionales de los datos obtenidos al dar a los participantes del experimento ideas adicionales sobre qué está pasando, al dar a los individuos la oportunidad de generar una impresión favorable, etc.

Los instrumentos de medición pueden ser reactivos en el sentido de que pueden cambiar el fenómeno que está siendo medido. Por ejemplo, el uso de cámaras en la presencia de experimentadores puede provocar que los individuos que están siendo estudiados se comporten atípicamente. La exposición a los instrumentos de medición en un pretest pueden sensibilizar a los individuos y afectar sus puntajes luego del test. Aún el tiempo de la medición puede provocar resultados engañosos, en tanto que el investigador puede medir los efectos de las variables independientes antes de que tengan tiempo de afectar a la variable dependiente o después de que sus efectos se hayan esfumado, por lo tanto, ocultando su efecto real. C. Hoveand y sus coautores descubrieron en su estudio pionero, que hablantes no estimados pueden no tener efecto inmediato en sus oyentes, pero pueden tener un efecto significativo un mes después, salvo que los oyentes recuerden la fuente.

Registrando las observaciones

Las observaciones en el laboratorio se registran en el lugar durante la sesión experimental. Luego, se asignan las unidades de observación a un sistema de categorías bien estructurado. La categorización también puede tener lugar durante la sesión experimental si el sistema de registro está preparado y pretestado con anticipación. Con un buen sistema de registro y observadores entrenados, el grado de inferencia requerido del observador es mínimo.

Experimentaciones de Campo

La principal diferencia entre experimentación de laboratorio y experimentación en el campo es, como el término indica, el escenario. La experimentación de laboratorio implica la introducción de condiciones que simulan ciertas características en un ambiente controlado. La experiencia en el campo, en cambio, consiste en un estudio en una situación natural donde son manipuladas una o más variables independientes por el investigador bajo situaciones lo más controladas posibles. En términos de los diseños de investigación, el contraste entre el experimento de laboratorio y el estudio

de campo no es tan fuerte. Sin embargo, las dificultades involucradas en el control de los factores intrínsecos y extrínsecos son considerablemente más fuertes en los experimentos de campo.

El mayor atractivo de los experimentos de campo es que permite al investigador analizar interacciones complejas, procesos y cambios en un escenario natural. La mayor debilidad es que las fuentes intrínsecas y extrínsecas de validez no pueden ser controladas de modo tan sistemático como en los experimentos de laboratorio. La autoselección de los participantes es un fuerte problema, la “randomización” es casi imposible. Las investigaciones piloto suelen ser más útiles en dichas situaciones. El investigador tiene que convencer (“make a convincing case”) de que los participantes en la investigación no tienden a diferenciarse en otro tipo de factores distintos al proceso casual de interés.

Existen problemas éticos en la experiencia de campo ¿Es ético exponer espectadores ante alguien que colapsa y simula estar gravemente enfermo? En los experimentos de laboratorio, los derechos de los participantes son protegidos por consentimiento informado y por un interrogatorio previo; los participantes son conscientes de formar parte de una investigación. Aunque los participantes accedan a la sesión experimental sin información sobre el objetivo del experimento, saben que serán informados cuando acabe la sesión. En los experimentos de campo, los individuos generalmente no son conscientes de que están participando en una investigación. En estas situaciones, el investigador tiene que asegurar que la privacidad de los individuos investigados no sea violada y de que serán protegidos de una vergüenza o estrés indebidos.

CAPÍTULO 10

Investigación por encuesta

En este capítulo exploramos el diseño de encuestas, considerando en detalle tres tipos de ellas: encuestas por correo, encuestas personales y encuestas telefónicas.

Los métodos observacionales de recolección de datos son apropiados para investigar fenómenos que pueden ser observados directamente por el investigador. Sin embargo, no todos los fenómenos son accesibles a la observación directa del investigador; frecuentemente, por lo tanto, el investigador debe recolectar datos pidiéndole a las personas que han experimentado cierto fenómeno que lo reconstruyan. El investigador aborda una muestra de individuos que presume han tenido ciertas experiencias e interroga a los mismos sobre esas experiencias. Las respuestas obtenidas constituyen los datos con los cuales son evaluadas las hipótesis de investigación. Para obtener información de los encuestados se utilizan principalmente tres métodos de investigación por encuesta: la encuesta por correo, la encuesta personal y la encuesta telefónica.

Encuesta por correo

La **encuesta por correo** es un método de encuesta impersonal. Bajo ciertas condiciones y para ciertos propósitos de investigación puede ser útil un método impersonal de recolección de datos. Vamos a examinar las ventajas y desventajas de este método.

Ventajas de la encuesta por correo

1. *Bajo costo.* La economía es uno de los atractivos más evidentes de la encuesta por correo. La misma no requiere un equipo de encuestadores entrenados; su realización involucra el costo de planificación, muestreo, imprenta, envío postal, y la provisión de sobres estampillados y con la dirección de respuesta. El procesamiento y análisis son usualmente más simples y económicos que los de otros métodos de investigación. El bajo costo en la administración de una encuesta por correo es particularmente evidente cuando la población bajo estudio está dispersa en áreas geográficas amplias. Bajo tales circunstancias, el costo de las encuestas personales puede ser prohibitivo, y la encuesta por correo puede ser el único instrumento practicable.

2. *Reducción del sesgo.* La encuesta por correo reduce el sesgo que podría resultar de las características personales del encuestador y la variabilidad de sus habilidades. Las situaciones de encuesta personal están llenas de oportunidades para el sesgo debido a la naturaleza de la interacción entre el encuestador y el encuestado. Esto puede ser evitado por completo con una encuesta por correo.

3. *Mayor anonimato.* La ausencia de un encuestador provee también un mayor anonimato. La garantía de anonimato en una encuesta por correo es especialmente útil cuando la encuesta toca temas sensibles, como el comportamiento sexual o el abuso infantil. En tales cuestiones, una encuesta por correo puede llevar a un mayor porcentaje de respuestas que una encuesta personal.

4. *Respuestas elaboradas y consultas.* La encuesta por correo también es preferible cuando las preguntas demandan una respuesta elaborada -antes que inmediata- o si las respuestas requieren la consulta de documentos personales o de otras personas.

5. *Accesibilidad.* Finalmente, la encuesta por correo permite una amplia cobertura geográfica a un costo mínimo. Por ejemplo, cuando una encuesta requiere una amplia cobertura y está dirigida a una población dispersa geográficamente, la encuesta personal puede acarrear grandes costos de transporte y una gran inversión de tiempo.

Desventajas de la encuesta por correo

1. *Requiere preguntas simples.* La encuesta por correo puede ser usada como un instrumento de recolección de datos sólo cuando las preguntas son lo suficientemente claras como para ser comprendidas sin asistencia, sobre la base de instrucciones y definiciones impresas.

2. *No es posible explorar (probing).* Las respuestas tienen que ser aceptadas como finales; no hay oportunidad de explorar más allá de la respuesta dada para clarificar respuestas ambiguas, o para evaluar el comportamiento no verbal de los encuestados.

3. *No hay control sobre quién llena el cuestionario.* En una encuesta por correo, los investigadores no tienen control sobre el entorno de los encuestados, y en consecuencia no pueden estar seguros de que la persona adecuada completa el cuestionario. Este puede ser respondido por una persona distinta del supuesto encuestado.

4. *Bajo porcentaje de respuestas.* La desventaja final de la encuesta por correo -y quizás su más serio problema- es que frecuentemente es difícil obtener un porcentaje de respuestas adecuado. En muchas encuestas por correo, los porcentajes de respuestas reportados son mucho más bajos que los de las encuestas personales. El porcentaje de respuesta típico para una encuesta personal es del 95 %, mientras que para una encuesta por correo sin seguimiento está entre el 20 y el 40 %. Los investigadores que utilizan encuestas por correo deben enfrentar casi siempre el problema de cómo estimar el efecto que la no-respuesta puede tener en sus hallazgos. Quienes no responden son con frecuencia bastante distintos de aquellos que contestan el cuestionario. Frecuentemente son los menos educados quienes no pueden responder, o los más móviles quienes no pueden ser localizados. En consecuencia, el grupo de encuestados probablemente no constituye el grupo representativo originalmente definido por los investigadores, y esto sin dudas introducirá un sesgo en el estudio.

Factores que afectan el porcentaje de respuestas en la encuesta por correo

La dificultad para asegurar un porcentaje de respuestas aceptable en una encuesta por correo requiere del uso de distintas estrategias que pueden ser llevadas a cabo para aumentar tal porcentaje. Vamos a revisar esas estrategias.

Auspicio (Sponsorship). El respaldo de una encuesta tiene un efecto significativo en cuanto a la motivación del encuestado para completarla y enviarla de vuelta. En ese sentido, la información sobre el auspiciante debe ser incluida, principalmente en la carta de presentación que acompaña al cuestionario. El auspiciante afecta el porcentaje de respuestas convenciendo al encuestado de la legitimidad y el valor del estudio, así también como la percepción de sanciones por una falla en la respuesta. Por ejemplo, la oficina estatal encargada del Censo tiene éxito al obtener cerca del 95 % de respuestas en su Encuesta Nacional de Salud. En el otro extremo, en algunas encuestas por correo sólo responde el 5 % de la muestra¹. En general, las encuestas auspiciadas por el gobierno obtienen una gran cantidad de respuestas mientras que las organizaciones escasamente conocidas obtienen pocas.

Estímulo. Los investigadores deben atraer a los encuestados y persuadirlos de que deben participar completando el cuestionario y enviándolo a vuelta de correo. Varios métodos pueden ser utilizados, con distintos grados de eficacia. Uno de ellos consiste en apelar a las buenas intenciones de los encuestados, diciéndoles que los investigadores necesitan su ayuda. Por ejemplo, un estudiante que lleva a cabo una encuesta para un trabajo escolar puede decir que su nota puede ser afectada por la respuesta al cuestionario².

Otro método ampliamente utilizado consiste en ofrecer una gratificación al encuestado, como un premio o una suma simbólica de dinero. El problema del ofrecimiento de dinero es que algunos encuestados se indignarán porque los investigadores consideran que el tiempo del encuestado vale tan poco, y de este modo pueden no responder absolutamente nada. Sin embargo, usualmente la gratificación es vista como un gesto simbólico, y los encuestados cooperan porque consideran que el estudio es valioso.

Quizás la estrategia más efectiva es apelar a los sentimientos altruistas de los encuestados y convencerlos de la importancia del estudio. En el siguiente ejemplo, la importancia del estudio y la potencial contribución del encuestado a su éxito están expresadas en la carta de presentación que acompaña al cuestionario.

Como ud. sabe, el servicio público de empleo es una parte importante de la estrategia federal, estadual y local para superar los problemas de empleo e ingresos de la gente económicamente desfavorecida y desempleada. No hay dudas de que el programa es necesario a lo largo del país... Usted probablemente sabe... que los programas del servicio público de empleo son bastante controvertidos y que su futuro estaría en peligro. Parte de las razones por las cuales estos programas son controvertidos es que no hay una evaluación sistemática de los beneficios de estos programas para los individuos empleados y las comunidades en las cuales se han llevado a cabo.

* Traducción de M. Teresa Almendros para la Cátedra Ruth Sautu.

¹ Floyd J. Fowler, Jr., *Survey Research Methods* (Newbury Park, Calif.:Sage, 1989), p.48.

² Kenneth D. Bailey, *Methods of Social Research* (New York: Free Press, 1987), p.156.

Porque esta evaluación específica tiene implicancias nacionales significativas, solicito encarecidamente dar al cuestionario adjunto una pronta atención, y agradezco su cooperación en esta evaluación.³

Formato del cuestionario y métodos de envío por correo. Hay varias cuestiones a considerar en el diseño de una encuesta por correo: la tipografía, el color, y la longitud y tipo de carta de presentación. Una inversión levemente mayor en formato y tipografía (p.ej. papel de alta calidad y espaciado adecuado) redundará en un más alto porcentaje de respuestas. Nótese que el uso de colores inusuales no es recomendable porque puede tener efectos negativos.⁴

Carta de presentación. Otro factor a ser considerado en el diseño del cuestionario es la carta de presentación. La carta de presentación debe lograr convencer a los encuestados de llenar el cuestionario y devolverlo por correo. Debe por lo tanto identificar al auspiciante del estudio, explicar los propósitos de la investigación, decir a los encuestados por qué es importante que llenen el cuestionario, y asegurarles que las respuestas serán mantenidas en estricta confidencialidad. El investigador debe elegir entre un carta formal o semi-personal. Se ha demostrado que una carta semi-personal genera un porcentaje de respuestas levemente mayor que una carta formal.

Tipo de envío postal. Una tema importante a considerar es el tipo de envío postal a utilizar. Los cuestionarios que no están acompañados por un sobre con la respuesta postal paga obtienen pocas respuestas. No es razonable esperar que el encuestado no sólo llene el cuestionario sino que consiga un sobre y luego vaya al correo a que lo pesen y estampillen. En consecuencia es una práctica común adjuntar un sobre estampillado y con la dirección de respuesta.

El método de diseño total (MDT). En años recientes, la recolección de datos a través de encuestas por correo ha mejorado considerablemente con la aplicación del método de diseño total, un conjunto estandarizado de procedimientos paso-a-paso⁵ que está dividido en dos partes: la construcción del cuestionario y la implementación de la encuesta.

Los principios seguidos en la construcción de los cuestionarios MDT incluyen una particular atención a los detalles tales como la cubierta del sobre que contiene el cuestionario, la presentación del cuestionario, y el orden de las preguntas. El MDT trata de asegurarse que el cuestionario será inmediatamente diferenciado de la publicidad enviada por correo.

El enfoque MDT de la implementación se ocupa principalmente de los procedimientos de seguimiento. La estrategia de seguimiento más común es enviar una carta recordatoria una semana después del primer envío a los encuestados que aún no han respondido. El segundo seguimiento consiste en el envío de otra carta recordatoria y un cuestionario de reemplazo con un sobre para la respuesta enviado a la tercera semana. Luego de siete semanas, se envía otra carta con un cuestionario de reemplazo,

³ Mickey L. Burnim, *An Evaluation of the Public Service Employment Projects in Florida Created under Title VI of the Comprehensive Employment and Training Act of 1973* (Tallahassee: Florida Department of Community Affairs, 1978), p.164.

⁴ Pamela L. Alreck and Robert B. Settle, *The Survey Research Handbook* (Homewood, Ill.:Irwin, 1985).

⁵ Donald A. Dillman, "Mail and Other Self-administered Questionnaires," in *Handbook of Survey Research*, ed. Peter H. Rossi, James D. Wright, and Andy B. Anderson (Orlando, Fla.: Academic Press, 1983), and Anton J. Nederhof, "Effects of a Final Telephone Reminder and Questionnaire Cover Design in Mail Surveys," *Social Science Research*, 17 (1988):353-361.

Tabla 10.1		
Porcentajes acumulados de respuestas (promedio)		
Tipo de envío postal	Tiempo	Porcentaje de respuestas promedio
1. Primer envío	Semana 1	23.8%
2. Tarjeta recordatoria	Semana 2	42.4%
3. Primer cuestionario de reemplazo	Semana 4	59.0%
4. Segundo reemplazo (envío certificado)	Semana 7	72.4%

Basada en Donald A. Dillman, James A. Christensen, Edward H. Carpenter, y Ralph M. Brooks, "Increasing Mail Questionnaire Response: A Four-State Comparison", *American Sociological Review*, 39 (1974): 755, y Donald A. Dillman y D.E. Moore, "Improving Response Rates to Mail Surveys: Results from Five Surveys," paper presentado en el encuentro anual de la American Association for Public Opinion Research, Hershey, Pa., 1983.

preferiblemente una carta certificada, a todos aquellos que no hayan respondido hasta el momento.

La efectividad de esos métodos de seguimiento fue testeada en grandes muestras estatales de la población general en cuatro estados. La tabla 10.1 muestra los porcentajes de respuestas promedio de los cuatro envíos postales usados en el estudio. Los resultados obtenidos por este estudio revelan la importancia de un seguimiento multietápico. Obsérvese que la etapa final incrementó el porcentaje de respuestas en más de un 13 %. Más aún, "con una metodología de envío postal disponible que proveerá sostenidamente una gran cantidad de respuestas, un bajo porcentaje de respuestas no puede ser disculpado más de lo que pueden serlo una inadecuada teoría o estadísticas inapropiadas"⁶. Recientemente se ha sugerido que el uso de correo certificado tiene algunas desventajas importantes⁷. El requerimiento de una firma para la recepción puede parecer coercitivo, y el costo en tiempo y dinero puede ser mayor si el encuestado tiene que ir al correo a retirar y devolver el cuestionario. El seguimiento final por correo certificado puede ser reemplazado por una llamada recordatoria, que es tan efectiva como el correo certificado para reducir la no respuesta.

Aunque el seguimiento es claramente un mecanismo importante para incrementar el porcentaje de respuestas, tiene algunos problemas. Primero, dado que las cartas de seguimiento y los cuestionarios son enviados solamente a los encuestados que no han respondido, es necesario identificar a todos los encuestados; por lo tanto no puede ser mantenido el anonimato. Una manera de superar esta dificultad es asegurar a los encuestados que las respuestas serán mantenidas en estricta confidencialidad. Otra limitación del seguimiento consiste en que la calidad del porcentaje de respuestas decae con los envíos postales sucesivos. Los encuestados que no responden en un primer momento pueden estar menos predispuestos a tomar seriamente el estudio y por lo tanto enviar un cuestionario incompleto, o sus respuestas pueden no ser confiables. El sesgo debido a este motivo puede ser estudiado comparando aquellos que respondieron inmediatamente con aquellos que respondieron luego de una o más etapas de seguimiento⁸.

Selección de los encuestados. La selección de los encuestados está determinada en gran parte por la naturaleza del estudio y las características de la población. Por lo tanto más allá de la definición de la población muestral, es muy poco

⁶ Donald A. Dillman, James A. Christensen, Edward H. Carpenter, and Ralph M. Brooks, "Increasing Mail Questionnaire Response: A Four-State Comparison." *American Sociological Review*, 39 (1974): 755.

⁷ Nederhof, "Effects," p.354.

⁸ Fowler, *Survey Research Methods*, p.54.

lo que puede hacerse en el proceso de selección para incrementar el porcentaje de respuestas. Sin embargo, reconocer que ciertas características de los encuestados están asociadas con un alto o bajo porcentaje de respuestas ayudará a determinar si se usará una encuesta por correo

Tabla 10.2		
Técnicas para incrementar el porcentaje de respuestas		
Método	Puntaje (de alto a bajo)	Condiciones Optimas
Seguimiento	1	Más de una estrategia de seguimiento. El seguimiento puede realizarse también por teléfono.
Estímulo	2	Las encuestas que otorgan una suma simbólica de dinero producen mejores resultados. Sin embargo, la población y el tipo de encuesta deben ser tenidos en cuenta.
Auspicio	3	Un auspiciante conocido produce los mejores resultados.
Carta de presentación	4	Apelar al altruismo parece producir los mejores resultados.
Formato	-	Una presentación estéticamente atractiva, un título que despierte interés, un formato de página atractivo.
Selección de los encuestados	-	<ul style="list-style-type: none"> ▪ Quienes no leen y no escriben son excluidos de la participación. ▪ El interés y la familiaridad con el tópico bajo investigación es un factor principal en la determinación del porcentaje de respuestas. ▪ Los mejor educados están mejor predispuestos a completar los cuestionarios. ▪ Los profesionales están mejor predispuestos a completar los cuestionarios.

Basado en Delbert C. Miller, *Handbook of Research Design and Social Measurement* (New York: McKay, 1977), pp.77-78; Pamela L. Alreck and Robert B. Settle, *The Survey Research Handbook* (Homewood, Ill.: Irwin, 1985), and Anton J. Nederhof, "Effects of a Final Telephone Reminder and Questionnaire Cover Design in Mail Surveys", *Social Science Research*, 17 (1988):353-361.

desde el comienzo o si deberá utilizarse otra estrategia para incrementar el porcentaje de respuestas. El aspecto más significativo para seleccionar a los encuestados es ver si constituyen un grupo heterogéneo u homogéneo. Un grupo heterogéneo podría estar compuesto por individuos con distintas procedencias étnicas y raciales, diferentes niveles de ingreso, o procedencia urbana y rural. Los grupos homogéneos, por el contrario, están constituidos por individuos con características similares. Los grupos heterogéneos son típicamente utilizados en encuestas de opinión, mientras que en estudios más especializados se envían cuestionarios a grupos más selectos, por ejemplo, a médicos, legisladores, autoridades municipales, profesores universitarios, o miembros de la cámara local de comercio. El porcentaje de respuestas para grupos selectos es usualmente más alto que el de la población general porque esos grupos están más predispuestos a identificarse con los objetivos del estudio y por lo tanto estarán más motivados a responder. Más allá de esta distinción, ciertas características de procedencia están asociadas con diferencias en los porcentajes de respuestas. Los encuestados más educados están más predispuestos a llenar y devolver los cuestionarios. El interés o la familiaridad con el tema en estudio es otro factor importante en la determinación del porcentaje de respuestas. Finalmente, en general, los profesionales tienden a tener los más altos porcentajes de respuesta entre todas las ocupaciones.

La tabla 10.2 ordena los distintos procedimientos discutidos hasta ahora de acuerdo a su eficacia relativa para incrementar el porcentaje de respuesta. Los puntajes fueron obtenidos sobre la base de distintos estudios que estiman el posible incremento del total de respuestas para cada procedimiento. No pudieron obtenerse puntajes para los últimos tres procedimientos.

Evaluando el porcentaje de respuestas

¿Qué porcentaje de respuestas es aceptable para una encuesta por correo? La mayoría de los investigadores intentan maximizar el porcentaje de respuestas usando alguna de las estrategias recién discutidas. Aun a pesar de esos esfuerzos, las encuestas por correo obtienen un porcentaje de respuestas no mayor al 50%. La no respuesta es un serio problema porque quienes no responden difieren considerablemente quienes que sí lo hacen. Por ejemplo, se ha demostrado que las encuestas por correo dirigidas a la población general dan por resultado un sesgo hacia arriba en educación: la gente más educada está más predispuesta a responder más rápido a las encuestas por correo. El sesgo resultante de la no respuesta puede limitar la posibilidad de hacer generalizaciones a toda la población.

La pregunta sobre qué constituye un porcentaje de respuestas aceptable no puede ser fácilmente respondida porque no hay un *standard* acordado para un porcentaje de respuestas mínimo. Por ejemplo, de las encuestas hechas bajo contrato con el gobierno federal se espera que obtengan un porcentaje de respuestas superior al 75%. Pero mientras que las organizaciones académicas de encuestas son normalmente capaces de alcanzar ese nivel, el porcentaje de respuestas a encuestas dirigidas por organizaciones más oscuras es considerablemente menor.

Finalmente, hay alguna evidencia de que los porcentajes de respuesta a encuestas por correo han ido mejorando con una mayor estandarización de las técnicas de seguimiento⁹.

Más aun, en años recientes, la investigación por encuesta se ha convertido en una herramienta ampliamente usada, no sólo por organizaciones de investigación y marketing sino también por el gobierno nacional y local. Algunos ciudadanos, aunque valoren los objetivos de la investigación, pueden encontrarse a sí mismos afrontando una decisión unilateral con respecto a cuáles y cuántos de los cuestionarios que reciben cada año deben responder.

Encuesta Personal

La encuesta personal es una situación interpersonal de rol cara a cara en la cual el encuestador hace a los encuestados preguntas diseñadas para obtener respuestas pertinentes con respecto a las hipótesis de investigación. Las preguntas, su redacción, y su secuencia definen la estructura de la encuesta.

La encuesta mediante un cuestionario estructurado

La forma más estructurada es la encuesta mediante un cuestionario estructurado, en la cual las preguntas, su redacción, y su secuencia están fijas y son idénticas para cada encuestado. Esto se hace para asegurarse de que cualquier variación entre las respuestas puede ser atribuida a las diferencias reales entre encuestados y no a variaciones de la encuesta. El investigador intenta reducir el riesgo

⁹ Nederhof, "Effects," p.356.

de que cambios en la redacción de las preguntas, por ejemplo, puedan provocar diferencias en las respuestas. La encuesta mediante un cuestionario estructurado está basada en tres presupuestos cruciales:

1. Que para cualquier objetivo de investigación “los encuestados tienen suficiente vocabulario en común de modo que es posible formular preguntas que tengan el mismo significado para cada uno de ellos”¹⁰.
2. Que es posible expresar todas las preguntas de modo que posean el mismo significado para cada encuestado.
3. Que si el significado de cada pregunta debe ser idéntico para cada encuestado, su contexto debe ser idéntico y, dado que todas las preguntas precedentes constituyen parte del contexto, la secuencia de las preguntas debe ser idéntica.

La entrevista focalizada

La segunda forma básica es la encuesta sin cuestionario estructurado o entrevista focalizada. Esta forma tiene cuatro características¹¹:

1. Tiene lugar con entrevistados que se sabe han estado involucrados en una experiencia particular.
2. Se refiere a situaciones que han sido analizadas con anterioridad a la entrevista.
3. Se lleva a cabo sobre la base de una guía de entrevista que especifica temas relacionados con las hipótesis de investigación.
4. Está focalizada en la experiencia de los sujetos de acuerdo a las situaciones bajo estudio.

Aunque el encuentro entre el entrevistador y el entrevistado está estructurado y los principales aspectos del estudio son explicados, se les da una libertad considerable a los entrevistados para expresar su propia definición de una situación que les es presentada. Por ejemplo, en su estudio sobre el matrimonio¹², Stacey Oliner empleó una entrevista focalizada que era “lo suficientemente maleable como para seguir orientaciones emergentes y lo suficientemente estandarizada como para registrar patrones fuertes”. La entrevista focalizada permite al investigador obtener detalles de las reacciones personales, emociones específicas, y temas similares. El entrevistador, habiendo estudiado previamente la situación, está alerta y sensitivo a las inconsistencias y omisiones de los datos que pueden ser necesarios para clarificar el problema.

La entrevista no dirigida

La última forma de encuesta personal es la entrevista *no estructurada* o **entrevista no dirigida**. Aquí no se utiliza un conjunto previamente especificado de preguntas, ni las preguntas son formuladas en un orden específico. Más aun, no se utiliza ningún cuestionario. Con poca o ninguna dirección de parte del entrevistador, los entrevistados son alentados a relatar sus experiencias, a describir eventos cualesquiera que sean significativos para ellos, a proveer sus propias definiciones de

¹⁰ Stephen Richardson, Barbara S. Dohrenwend, and David Klein, *Interviewing: Its Forms and Functions* (New York: Basic Books, 1965), p. 40.

¹¹ Robert K. Merton and Patricia L. Kendal, “The focused Interview,” *American Journal of Sociology*, 51 (1946): 541-557.

¹² Stacey J. Oliner, *Best Friends and Marriage* (Berkeley: University of California Press, (1989), p.xvi.

sus propias situaciones, y a revelar sus opiniones y actitudes tal como ellos las ven. El entrevistador tiene alto grado de libertad para explorar distintas áreas y para hacer preguntas específicas durante el transcurso de la entrevista. Por ejemplo, el estudio de Eleanor Miller sobre prostitutas está basado en entrevistas no dirigidas¹³:

Setenta mujeres aceptaron hacer entrevistas grabadas en las cuales compartieron conmigo los detalles de sus vidas. Se prestó especial atención en la iniciación de esas mujeres en la prostitución y el desarrollo de una carrera como prostituta. Aunque los mismos temas amplios fueron introducidos en cada entrevista, muchas de mis preguntas cambiaron a través del tiempo. Las primeras entrevistas grabadas fueron reproducidas una y otra vez antes de ser registradas. De esas horas de audición surgieron hipótesis tentativas y categorías emergentes de comportamiento. Durante las siguientes entrevistas, ... pude introducir preguntas para testear las hipótesis tentativas.

Las diferencias en los estilos de encuestar en los tres tipos de encuestas están ilustradas en los cuadros 10.1, 10.2 y 10.3, concernientes al mismo problema de investigación. El propósito del estudio es descubrir los tipos de conflicto entre padres y adolescentes y su relación con el crimen juvenil. Las encuestas son llevadas a cabo con dos grupos de adolescentes. Uno está compuesto por adolescentes que no han cometido delitos, y el segundo por adolescentes que se sabe han cometido diversos delitos juveniles.

Una encuesta personal puede ser completamente estructurada o no estructurada, según se muestra en los cuadros. Alternativamente, una encuesta personal puede combinar elementos estructurados y no estructurados, dependiendo del propósito del estudio. Por ejemplo, un investigador puede usar la encuesta mediante un cuestionario estructurado para la mayoría de las preguntas pero elegir el formato no dirigido para preguntas que son particularmente sensibles.

Encuesta personal versus encuesta por correo

Ventajas de la encuesta personal

1. *Flexibilidad.* La encuesta personal permite una gran flexibilidad en el proceso de interrogación, y a mayor flexibilidad, menos estructurada es la encuesta. La encuesta personal permite al encuestador determinar la redacción de las preguntas, clarificar términos que no están claros, controlar el orden el cual las preguntas son efectuadas, e indagar en profundidad para obtener información y detalles adicionales.

2. *Control de la situación de encuesta.* Una ventaja principal de la encuesta personal es que permite un mayor control sobre la situación de encuesta. Un encuestador puede asegurarse de que los encuestados contestan las preguntas en el orden apropiado o que contestan ciertas preguntas antes de que se les hagan las subsiguientes. Mas aún, en una situación de encuesta personal es posible estandarizar el ambiente para asegurarse de que la encuesta se lleva a cabo en privado; de este modo los encuestados no tendrán la oportunidad de consultarse antes de dar sus respuestas. También es posible registrar el tiempo y lugar exactos de la encuesta; esto permite al investigador interpretar las respuestas con más precisión, especialmente en casos en los cuales un evento ocurrido sobre la fecha de la encuesta podría haber influido las respuestas de los encuestados.

¹³ Eleanor M. Miller, *Street Woman* (Philadelphia: Temple University Press, 1986).

3. *Alto porcentaje de respuestas.* La encuesta personal da por resultado un más alto porcentaje de respuesta que la encuesta por correo. En una encuesta personal pueden ser fácilmente obtenidas las respuestas de encuestados que normalmente no responden a una encuesta por correo. Esto es igualmente cierto para personas que tienen dificultades para leer o escribir, que no entienden completamente el idioma, o que simplemente no desean tomarse el tiempo de escribir sus respuestas y devolver el cuestionario por correo.

2. *Recolección de información suplementaria.* Un encuestador puede recolectar información suplementaria sobre los encuestados. Esta puede incluir información de base sobre las características personales de los encuestados y su entorno que puede ayudar al investigador en la interpretación de los resultados. Más aun, una situación de encuesta personal a menudo capta reacciones espontáneas que el encuestador puede registrar y que puede ser útil en la etapa de análisis de datos.

Desventajas de la encuesta personal

1. *Mayor costo.* El costo de los estudios mediante encuestas personales es significativamente mayor que el realizado mediante encuestas por correo. Hay costos relativos a la selección, entrenamiento, y supervisión de los encuestadores, su paga, y en el viaje y tiempo requerido para realizar las encuestas. Más aun, el costo del registro y procesamiento de la información obtenida en encuestas no estructuradas es especialmente alto.

Cuadro 10.1**La encuesta mediante un cuestionario estructurado**

Explicación del encuestador al encuestado: Estamos interesados en los tipos de problemas que los adolescentes tienen con sus padres. Necesitamos saber cuántos adolescentes tienen conflictos con sus padres y cuáles son esos conflictos. Tenemos aquí una lista de algunos posibles conflictos. Piense en su propia situación, y marque qué conflictos ha tenido y cuán frecuentemente han sucedido. Asegúrese de poner una marca en cada fila. Si usted nunca ha tenido ese tipo de conflicto, marque la primera columna, donde dice "nunca". (Entregue al encuestado la primera tarjeta que trata de los conflictos en torno al uso del automóvil, diciendo, "Si no entiende alguna de las situaciones enumeradas o hay algo que quiera mencionar sobre el desacuerdo con sus padres sobre el automóvil, hágalo saber y hablaremos al respecto.")

Automóvil	Nunca	Sólo una vez	Más de una vez	Muchas veces
1. Querer aprender a manejar				
2. Querer obtener una licencia para conducir				
3. Querer usar el automóvil familiar				
4. Usar demasiado el automóvil				
5. Mantener limpio el automóvil				
6. Reparar el auto				
7. Manejar el auto de otro				
8. Querer un auto propio				
9. La manera en que se conduce el auto propio				
10. Otras				

(Cuando el encuestado complete todos los renglones, entregue la tarjeta número 2, diciendo, "esta es una lista de los tipos de conflictos que los adolescentes tienen con sus padres sobre sus amigos del mismo sexo. Haga lo mismo que hizo con la lista anterior")

Adaptado en base a Raymond L. Gorden, *Interviewing: Strategy, Techniques, and Tactics*, 2d ed. (Homewood, Ill.:Dorsey, 1975), pp.63-65

Cuadro 10.2

La entrevista focalizada

Instrucciones para el entrevistador: Su tarea es descubrir tantos tipos específicos de conflictos y tensiones entre los adolescentes y sus padres como sea posible. Cuanto más concreta y detallada sea la descripción de cada tipo de conflicto, mejor. Aunque hay cuatro áreas de posibles conflictos que queremos explorar (enumeradas en la pregunta 3 más abajo), ud. no debe mencionar ningún área hasta después de que haya hecho las primeras dos preguntas en el orden indicado. La primera pregunta tiene un abordaje indirecto, dándole tiempo para establecer una buena relación con el encuestado.

1. ¿Qué tipo de problemas tienen los adolescentes en su relación con sus padres?
(Posibles pruebas: Siempre están de acuerdo con sus padres? ¿Alguno de sus amigos tiene “padres problemáticos”?)
3. ¿Qué tipo de desacuerdos tiene ud. con sus padres? (Posibles pruebas: ¿Le causan algunos problemas? ¿De qué manera tratan de ponerle límites? ¿A ellos les gustan las mismas cosas que a ud.?)
4. Alguna vez ha tenido desacuerdos con alguno de sus padres sobre:
 - a. el uso del auto familiar
 - b. amigos del mismo sexo
 - c. noviazgo
 - d. fumar

Adaptado en base a Raymond L. Gorden, *Interviewing: Strategy, Techniques, and Tactics*, 2d ed. (Homewood, Ill.: Dorsey, 1975), pp.63-65.

Cuadro 10.3

La entrevista no dirigida

Instrucciones para el entrevistador: Descubra los tipos de conflictos que el adolescente ha tenido con sus padres. Los conflictos deben incluir los desacuerdos; las tensiones debidas a desacuerdos pasados, presentes o potenciales; discusiones abiertas; y conflictos físicos. Esté atento a captar la mayor cantidad de categorías y ejemplos de conflictos y tensiones que sea posible.

Adaptado en base a Raymond L. Gorden, *Interviewing: Strategy, Techniques, and Tactics*, 2d ed. (Homewood, Ill.: Dorsey, 1975), pp.63-65.

2. *Sesgo introducido por el encuestador.* La gran flexibilidad, que es la mayor ventaja de la encuesta personal, da lugar a la influencia y al sesgo del encuestador. La falta de

estandarización en el proceso de recolección de datos también hace a la encuesta personal altamente vulnerable al sesgo del encuestador. Aunque los encuestadores son instruidos para mantenerse objetivos y evitar comunicar puntos de vista personales, a menudo emiten signos que pueden influir en las respuestas de los encuestados. Aun cuando se evitan los signos verbales, la comunicación no verbal puede escapar al control del encuestador. A veces incluso la raza o género del encuestador puede influir en los encuestados, quienes en un intento de complacer al encuestador pueden dar respuestas socialmente admirables pero potencialmente falsas.

3. *Falta de anonimato.* La encuesta personal carece del anonimato que sí tiene la encuesta por correo. Frecuentemente el encuestador sabe todo o mucho de los potenciales encuestados (o al menos su nombre, dirección, y el número de teléfono). De este modo el encuestado puede sentirse amenazado o intimidado por el encuestador, especialmente si el tema o algunas preguntas tratan acerca de una cuestión delicada.

Principios de la encuesta personal

Ahora pasaremos a una discusión más detallada de los principios y procedimientos de la encuesta personal. El primer paso en el proceso de encuesta personal es conseguir que el encuestado coopere y provea la información deseada. Tres factores ayudan a motivar al encuestado a cooperar.

1. Los encuestados necesitan sentir que su interacción con el encuestador será placentera y satisfactoria. Es recomendable para los encuestadores presentarse a los encuestados como personas comprensivas y de fácil diálogo.
2. Los encuestados necesitan ver que el estudio es valioso. Los encuestados deben sentir no sólo que el estudio puede ser beneficioso para ellos sino también que trata de un tema significativo y que su cooperación es importante. Los encuestadores deben interesar a los encuestados en el estudio señalando su importancia y la contribución que éstos pueden hacer mediante su cooperación.
3. Es necesario que sean superadas las barreras hacia la encuesta en las concepciones del encuestado. Los encuestadores deben corregir las ideas erróneas. Algunos encuestados pueden sospechar de los encuestadores, viéndolos como vendedores o representantes del gobierno. Los encuestadores deben explicar, de una manera amable, el propósito del estudio, el método para seleccionar encuestados, y la naturaleza confidencial de la encuesta.

El centro de investigaciones por encuesta del Instituto de Investigación Social de la Universidad de Michigan provee algunos consejos útiles sobre cómo el encuestador debe presentarse al encuestado:

1. Diga al encuestado quién es usted y a quién representa.
2. Diga al encuestado qué está haciendo de un modo que estimule su interés.
3. Diga al encuestado cómo él o ella ha sido elegido.
4. Haga que las instrucciones preliminares sean breves.
5. Adapte su modo de establecer el contacto a la situación.
6. Trate de crear una relación de confianza y entendimiento entre usted y el encuestado.

Luego de la presentación inicial, el encuestador está listo para comenzar la encuesta. Hay técnicas específicas que el encuestador puede usar en este proceso:

1. El cuestionario debe ser seguido, pero puede ser utilizado informalmente.
2. La encuesta debe ser realizada en una atmósfera informal y relajada, y el encuestador debería evitar crear la impresión de que lo que está ocurriendo es un interrogatorio o un examen.
3. Las preguntas deben ser preguntadas exactamente como están redactadas en el cuestionario. Esto es particularmente importante; incluso para pequeños cambios en la forma en que el cuestionario es preguntado puede cambiar la respuesta obtenida. Varios estudios han mostrado que incluso pequeñas omisiones o cambios en la enunciación de las preguntas pueden distorsionar los resultados.
4. Las preguntas deben ser presentadas en el mismo orden que en el cuestionario. El orden de las preguntas ha sido planeado por el investigador para lograr una continuidad y asegurarse que las respuestas de los encuestados no serán influidas por sus respuestas a preguntas previas. Para estandarizar la encuesta cada encuestador debe seguir la misma secuencia que la señalada por el investigador.

5. Las preguntas que no son interpretadas o comprendidas correctamente deben ser repetidas y clarificadas. En la mayoría de los casos, los encuestados no tendrán problema en comprender o interpretar una pregunta. A lo sumo, algunas personas necesitarán más tiempo para responder una pregunta particular. Pero ocasionalmente, los encuestados que tengan problemas de lenguaje o audición tendrán dificultades en comprender una pregunta. El encuestador debe entonces repetir la pregunta. Sólo en raras ocasiones el encuestador debe reformular la pregunta, y sólo si está convencido de que de otro modo el encuestado podría malinterpretar la pregunta.

Exploración (Probing)

En el manual del encuestador del Centro de Investigación por encuesta de la Universidad de Michigan, la exploración es definida como

la técnica usada por el encuestador para estimular la discusión y obtener más información. Una pregunta ha sido formulada y se ha obtenido una respuesta. Por una serie de motivos, la respuesta puede ser inadecuada y requerir que el investigador busque más información para lograr los objetivos de investigación. La exploración (probing) es el acto de obtener esta información adicional.

Las técnicas de exploración tienen dos funciones principales: motivan al encuestado a elaborar o clarificar una respuesta o a explicar las razones que subyacen a la respuesta, y ayudan a focalizar la conversación en el tema específico de la encuesta.

En general, cuanto menos estructurada sea la encuesta, más importante es la exploración como un instrumento para obtener y propiciar más información.

Lo siguiente es una muestra de la exploración realizada por el entrevistador para obtener información adicional mediante la repetición de las afirmaciones de los encuestados sin incluir una pregunta directa.

R: El principal motivo por el que vine a esta facultad fue la combinación de altos estándares académicos y el programa de estudio. Me resultó muy atractivo.

I: ¿Le resultó muy atractivo?

R: Sí.

I: ¿Podría decirme más exactamente por qué tuvo ese atractivo para ud.?

R: No sé, fue el lugar que sonaba menos formal y anticuado que muchos lugares con un programa académico similar.

I: ¿No le gustan los lugares formales y anticuados?

R: Muchos lugares pasan la mayor parte del tiempo tratando de descubrir una forma de controlar a los estudiantes, asumiendo que ellos son completamente incapaces de auto-controlarse.

I: ¿Por qué supone que esta facultad tiene menos supervisión por parte de la administración?

R: Bueno, es parte de su filosofía educativa...

I: A ver si entendí correctamente: a ud. le gusta una facultad con altos estándares académicos, pero una que no sea demasiado anticuada y que opere sobre el presupuesto de que los alumnos pueden ejercitar su autocontrol.

R: Exacto.

Encuesta telefónica

La encuesta telefónica puede ser caracterizada como un método semipersonal de recolección de información. No hace mucho, las encuestas telefónicas eran vistas con escepticismo o total desconfianza. Algunos textos explícitamente aconsejan a sus

lectores que las eviten. La principal razón de la resistencia a usar la encuesta telefónica fueron sus grandes posibilidades de introducir un serio sesgo de muestreo. Cuando una proporción sustancial de la población no tenía acceso a los teléfonos, la muestra tendía a sobre representar a aquellos que tenían un buen pasar económico y podían costearse un teléfono. Más recientemente, sin embargo, las encuestas telefónicas han ganado la aceptación general como un método legítimo de recolección de datos en las ciencias sociales.

La principal razón para emplear encuestas telefónicas más extensivamente es que hoy en día se estima que tienen una cobertura de más de nueve décimos de la población. En 1958, sólo el 72,5 % de los hogares estadounidenses tenían acceso al teléfono; hacia el fin de los 80 la cifra estaba cerca del 98 %. Además, las presiones económicas han hecho a las encuestas telefónicas más atractivas. El aumento de los salarios y los costos del combustible han hecho a la encuesta personal extremadamente cara. En comparación, el teléfono es conveniente, y produce un ahorro muy significativo. Más aun, la encuesta telefónica redundante en un mayor porcentaje de respuestas que la encuesta personal. En algunas áreas metropolitanas, a la gente le incomoda abrir la puerta a desconocidos. También se ha vuelto crecientemente difícil encontrar en su casa a los encuestados, dado el aumento de la participación de las mujeres casadas en la fuerza de trabajo.

Los cambios tecnológicos y los adelantos en equipamiento telefónico han hecho más sencilla la encuesta telefónica. Se ha hecho posible extraer una muestra aleatoria de números de teléfono mediante un proceso llamado discado aleatorio de dígitos (random-digit dialing, RDD). Pero más allá de las obvias ventajas en costo y velocidad que da la encuesta telefónica, queda en pie la pregunta sobre si las encuestas telefónicas son una alternativa a la encuesta cara a cara. En el principal experimento diseñado para responder esta pregunta, William Klecka y Alfred Tuchfarber replicaron una extensa investigación a través de encuestas personales mediante una encuesta telefónica RDD. La encuesta personal sobre victimización fue dirigida por la oficina estadounidense del censo en 1974. Las dos muestras fueron comparadas en cuanto a medidas de características demográficas de la victimización y actitudes hacia el crimen y la policía. Los resultados fueron muy similares, indicando que el discado de dígitos aleatorios es una alternativa a la encuesta personal precisa y efectiva en cuanto a costos. Estudios más recientes que comparan respuestas a las mismas preguntas en encuestas por correo, telefónicas y personales encontraron diferencias pequeñas en su validez.¹

Más allá de su relativa precisión, la encuesta telefónica tiende a incrementar la calidad del dato. En la mayoría de los casos, los encuestadores telefónicos están trabajando en una oficina central, y su trabajo puede ser monitoreado constantemente por el personal de supervisión. Esto ayuda a asegurar que las preguntas están siendo efectuadas correctamente y que los problemas pueden ser identificados inmediatamente y corregidos.

Uno de los últimos desarrollos en encuestas telefónicas es el uso de cuestionarios computarizados. En encuestas telefónicas asistidas por computadora (computer-assisted telephone interviewing, CATI), el encuestador se sienta ante la terminal de computadora y, cuando una pregunta parpadea en la pantalla, ésta es efectuada por teléfono. Las respuestas de los encuestados son tipeadas y codificadas directamente en un disco, y la pregunta siguiente aparece en la pantalla. Entre las ventajas del CATI se encuentran su velocidad y el uso de instrucciones complejas, programadas con anticipación. Sin embargo, CATI no es adecuada para preguntas abiertas.

Por otra parte, la debilidad del método no puede ser ignorada. La encuesta telefónica ha creado un nuevo tipo de no respuesta: la encuesta interrumpida. En alrededor un 4% de los llamados, los encuestados terminan la encuesta antes de que

¹ Seymour Sudman and Norman M. Bradburn, *Asking Questions* (San Francisco: Jossey-Bass, 1982).

ésta se complete, algo raro en encuestas personales. Las encuestas telefónicas producen también menos información; los encuestadores no pueden describir las características de los encuestados o su entorno en detalle. Más aun, proporcionalmente más encuestados telefónicos indican que se sienten incómodos al tratar ciertos temas telefónicamente, especialmente el estatus económico y las actitudes políticas.

En resumen, la encuesta telefónica debería ser usada como una alternativa a la encuesta personal bajo ciertas circunstancias, especialmente cuando el cuestionario es relativamente simple. Sin embargo, la pregunta sobre si las encuestas personal y telefónica son intercambiables permanece sin respuesta. En el futuro las encuestas podrían ser dirigidas totalmente por teléfono; otras podrían combinar entrevistas personales y telefónicas para que ambas puedan complementarse y dar mayor precisión y un porcentaje de respuestas incrementado.

Comparando los tres métodos de investigación

Para decidir cuál método es el más adecuado para la propia investigación, hay que evaluar cuál criterio es más significativo para el objetivo de investigación. Por ejemplo, si el investigador planea una larga encuesta con una muestra representativa de la población general y desea controlar el comportamiento no verbal, y si se dispone de fondos suficientes, es preferible una forma de encuesta personal.

Por el contrario, si la encuesta puede ser simplificada, y si los fondos y la velocidad son una preocupación, la encuesta telefónica puede ser usada para recolectar información. Si se utilizará un cuestionario bastante largo, o que incluye preguntas amenazantes o delicadas, y especialmente si la población a ser investigada está relativamente dispersa geográficamente o es una población selecta, la encuesta por correo puede ser considerada una alternativa.

Tabla 10.4
Evaluación de los tres métodos de encuesta

Criterio	Encuesta Personal	Correo	Teléfono
Costo	Alto	Bajo	Moderado
Porcentaje de respuestas	Alto	Bajo	Alto
Control de la situación de encuesta	Alto	Bajo	Moderado
Aplicabilidad a poblaciones geográficamente dispersas	Moderado	Alto	Moderado
Aplicabilidad a poblaciones heterogéneas	Alto	Bajo	Alto
Recolección de información detallada	Alto	Moderado	Moderado
Velocidad	Bajo	Bajo	Alto

Conclusión

La investigación por encuestas es uno de los más importantes métodos de recolección de datos en las ciencias sociales, y por ello es utilizado ampliamente para recolectar información sobre numerosos temas de investigación. En años recientes, con la demanda pública de transparencia, el énfasis en los instrumentos de encuesta se ha incrementado. Hay señales de que la investigación por encuesta se está convirtiendo en una herramienta ampliamente usada por distintas organizaciones gubernamentales. Estudios de los gobiernos locales indican que el 50% de las ciudades con poblaciones de más de 100.000 y distritos de más de 250.000 han usado alguna forma de encuesta. Con el crecimiento de la cantidad de encuestas llevadas a cabo, el método ha sido crecientemente criticado. Comentarios como “hacer las cosas bien en las ciencias sociales no es fácil”, “la muestra de encuestados potenciales fue una mezcla de varios procedimientos”, y “yo no confiaría en ninguna encuesta con un porcentaje de respuestas semejante”, son típicos. Aunque a veces esos señalamientos están justificados, frecuentemente no están basados en hechos o son simplemente críticas sin respaldo. No obstante, necesitamos un conjunto de criterios que nos ayuden a evaluar la utilidad de las encuestas, detectar y controlar errores en ellas, y quizás compensar los errores cuando sea posible.²

Hace medio siglo, Edward Deming escribió un artículo, actualmente un clásico, llamado “Sobre los errores en encuestas”³ En ese artículo, Deming detalla 13 errores potenciales que deben ser tenidos en cuenta cuando se planifica una encuesta y cuando se evalúa sus resultados. Los factores más importantes que podrían convertirse en errores en las encuestas fueron discutidos en este capítulo: sesgo del entrevistador, bajo porcentaje de respuestas, y dificultades al hacer preguntas sobre temas delicados. Reuben Cohen hizo las siguientes acotaciones teniendo en cuenta esos errores potenciales en su discurso presidencial en la Asociación Estadounidense para la Investigación de la Opinión Pública:

Hace alrededor de 30 años, me hicieron llegar una reimpresión de la lista de errores en encuestas de W. Edward Deming. El mensaje era muy obvio: ahora que sabe acerca de ellos, no los cometa. Con mi relativa inexperiencia y mi optimismo eterno, acepté el desafío. Mi primera intención fue tratar de hacer la encuesta perfecta. Estoy aún intentándolo, pero tendría que saber más. Descubrí rápidamente la ley de Murphy: si algo puede salir mal, probablemente saldrá mal. Pero también descubrí algo más. Aun sin las restricciones de tiempo y presupuesto de las cuales la mayoría de nosotros nos quejamos, no hay encuestas perfectas. Cada encuesta tiene sus imperfecciones. El mundo no está idealmente adecuado a nuestro trabajo. Lo mejor que podemos hacer es pensar detenidamente el abordaje ideal del diseño de la encuesta, de su implementación, o de su análisis -lo que haríamos si tuviéramos las posibilidades- luego acercarse tanto como sea posible al ideal dentro de las restricciones de tiempo y presupuesto que rigen nuestro trabajo.⁴

Y para los lectores que puedan estar desalentados por estos menos que perfectos objetivos, ofrecemos el siguiente consejo:

El trabajo práctico consiste en buena parte en adivinar qué irregularidades, dónde, y cuántas uno puede llegar a tolerar. Lo mismo es cierto para la investigación por encuestas. Debe estar bien hecha. Puede y debe conformarse, aún si no perfectamente, a un abordaje ideal.

² Gregory Daneke and Patricia Klobus Edwards, “Survey Research for Public Administrators”, *Public Administration Review*, 39 (1979):421-426.

³ W. Edward Deming, *Some Theory of Sampling* (New York: Wiley, 1950).

⁴ Reuben Cohen, “Close Enough for All Practical Purposes”, *Public Opinion Quarterly*, 43 (1979):421-422.

Resumen

1. En este capítulo, discutimos la encuesta como un método de recolección de datos. Fueron descriptos tres métodos: la encuesta por correo, la encuesta cara a cara, y la encuesta telefónica.

2. La encuesta por correo está considerada como un método de encuesta impersonal. Sus mayores ventajas son el bajo costo, el relativamente pequeño sesgo, el anonimato, y la accesibilidad. Sus desventajas son un bajo porcentaje de respuestas, la falta de oportunidad para la indagación en profundidad, y la falta de control sobre quién llena el cuestionario.

3. La dificultad para asegurarse un porcentaje de respuestas aceptable en la encuesta por correo lleva al uso de distintas estrategias que se sabe afectan el porcentaje de respuestas. Entre ellas, el más efectivo es el uso de correspondencia para realizar el seguimiento, el auspicio de la encuesta, y el atractivo del cuestionario. El formato del cuestionario y los métodos de envío postal utilizados afectarán también el porcentaje de respuesta.

4. La encuesta personal es una situación cara a cara en la cual un encuestador hace a los encuestados preguntas diseñadas para obtener respuestas pertinentes con respecto a las hipótesis de investigación. La forma más estructurada de encuesta es la encuesta mediante un cuestionario estructurado, en la cual las preguntas, su redacción, y su orden están fijos y son idénticos para cada encuestado. La entrevista focalizada sigue una guía de entrevista especificando temas relacionados con las hipótesis de investigación y da libertad considerable a los entrevistados para expresar sus puntos de vista. Finalmente, las entrevistas no dirigidas dan un amplio margen de libertad para explorar distintas áreas y realizar preguntas específicas durante el transcurso de la entrevista.

5. La encuesta telefónica ha ganado la aceptación general como un sustituto de la encuesta personal. La encuesta telefónica es conveniente y efectiva en costos. Además, a veces da por resultado un mayor porcentaje de respuestas que la encuesta personal. Mas aun, el cambio tecnológico y el avance del equipamiento telefónico hicieron a la encuesta telefónica más fácil, especialmente cuando se utiliza el discado aleatorio de dígitos y la encuesta telefónica asistida por computadora.

Conceptos claves

encuesta por correo
porcentaje de respuestas
seguimiento
encuesta mediante un cuestionario estructurado
entrevista focalizada
entrevista no dirigida
exploración
discado aleatorio de dígitos
encuesta telefónica asistida por computadora

Capítulo 11

Construcción de cuestionarios

En este capítulo discutiremos la formulación de tipos específicos de preguntas y formatos de preguntas en el diseño de encuestas. Se analizará en detalle la elección de las palabras en la construcción de preguntas, y se detallarán las trampas en las que se suele caer al construir un cuestionario.

En este capítulo nos enfocaremos en el cuestionario como el principal instrumento en la investigación por encuesta. Comenzaremos discutiendo el fundamento de todo cuestionario, la pregunta. Luego centraremos nuestra atención en el contenido de las preguntas y en la diferencia entre preguntas abiertas, cerradas y preguntas de contingencia, y analizaremos sus distintos formatos y secuencias. Luego analizaremos los posibles sesgos en la formulación de las preguntas. Finalmente, daremos importantes sugerencias acerca de la carta de presentación que acompaña el cuestionario y las instrucciones incluidas en él.

La pregunta

El fundamento de todo cuestionario es la **pregunta**. El cuestionario debe traducir los objetivos de investigación en preguntas específicas; las respuestas a estas preguntas proveerán los datos para probar las hipótesis. La pregunta debe también motivar al encuestado a proveer la información buscada. Las principales cuestiones a considerar en la formulación de las preguntas son: contenido, estructura, formato y secuencia.

Contenido de las preguntas

Las preguntas de una encuesta pueden involucrar hechos, opiniones, actitudes, las motivaciones del encuestado y su nivel de familiaridad con un cierto tema. La mayoría de las preguntas pueden ser clasificadas en una de dos categorías: preguntas fácticas y preguntas sobre experiencias subjetivas.

Preguntas fácticas:

Las **preguntas fácticas** están diseñadas para obtener información objetiva acerca del entorno, ambiente y hábitos del encuestado. El tipo más común son las preguntas acerca del entorno, que son utilizadas mayormente para proveer información en la que el encuestado pueda ser clasificado, como el género, edad, estado civil, educación o ingresos. Tales clasificaciones pueden ayudar a explicar diferencias en comportamientos y actitudes. El que sigue es un ejemplo de una pregunta de este tipo:

¿Cuál fue su último año escolar alcanzado? (por favor, marcar la respuesta que corresponda)

- Octavo grado o menor
- Noveno o décimo grado
- Onceavo o doceavo grado: graduado de escuela secundaria: _Sí _No
- Uno o dos años de universidad
- Tres o cuatro años de universidad: graduado universitario: _Sí _No
- Cinco o más años de universidad

Otros tipos de preguntas fácticas buscan proveer información acerca de el ambiente social del encuestado (“Podría por favor decirme, ¿quienes son las personas viviendo en su casa?”), sus formas de transporte (“¿Cómo va generalmente al trabajo?”), o las actividades que realiza en momentos de ocio (“¿Cuán seguido va al cine?”).

Se suele considerar a este tipo de preguntas como más fáciles de diseñar. Sin embargo, incluso las preguntas fácticas pueden presentar problemas para el investigador. Cuan certeramente la gente reporta depende de que se pregunta y como están siendo formuladas las preguntas. Existen cuatro tipos de razones por las cuales los encuestados pueden dar respuestas no precisas a este tipo de preguntas³²:

1. No tienen la información.
2. No pueden recordar la información.
3. No entienden la pregunta.
4. No quieren responderla.

El investigador puede tomar varios pasos para aumentar la precisión, incluyendo animar al encuestado a consultar con otros miembros de la familia, preguntando más de una pregunta acerca de una misma cuestión, y haciendo que el encuestado se sienta cómodo al preguntar acerca de eventos que puedan resultar embarazosos para ellos.

Preguntas sobre experiencias subjetivas:

Las experiencias subjetivas involucran las creencias, actitudes, sentimientos y opiniones de los encuestados.³³ Las preguntas acerca de las actitudes suelen aparecer en los estudios conducidos por las ciencias sociales. Las **actitudes** se definen como orientaciones generales que inclinan a una persona a actuar o reaccionar de cierto modo al ser confrontado por determinado estímulo. Aquí presentamos como ejemplo una pregunta sobre las actitudes acerca del aborto:

Sobre la cuestión del aborto, ¿cómo describiría su posición personal?

1. Favorable
2. Neutral
3. De oposición

Las actitudes de un individuo son expresadas, en el discurso o el comportamiento, solo cuando el objeto de esta actitud es percibido. Una persona puede tener fuertes actitudes en contra del aborto, pero ellas solo son incitadas y comunicadas cuando esa persona se encuentra con una cuestión conectada con el aborto, o al ser confrontada con un estímulo tal como una pregunta en una entrevista.

Las actitudes pueden ser descriptas por su contenido (de que se trata la actitud), su dirección (sentimientos positivos, neutrales o negativos acerca del objeto o tema en

³² Floyd J. Fowler, Jr., *Survey Research Methods* (Newbery Park, Calif.: Sage, 1989), p. 91.

³³ Royce Singleton, Jr., Bruce C. Straits, Margaret M. Straits y Ronald J. McAllister, *Approaches to Social Research* (Oxford: Oxford University Press, 1988), p. 272.

cuestión) y su intensidad (una actitud puede ser sostenida con mayor o menor vehemencia). Para una persona, el aborto puede tener un interés pasajero; para otra, puede ser de gran significación y llevar a dicha persona a unirse a una asociación anti aborto.

En general, estamos interesados en medir actitudes porque dan cuenta de las inclinaciones generales del entrevistado. El estudio de una opinión es solo interesante en tanto es símbolo de una actitud. La principal diferencia entre preguntar por opiniones y medir actitudes es que una **opinión** es generalmente medida estimando que proporción de una población encuestada dice estar de acuerdo con un enunciado. Las actitudes son medias mediante escalas de actitudes, consistentes de entre cinco y veinticuatro (o más) afirmaciones, mediante las cuales se le pide al entrevistado que exprese su acuerdo con ellas o no. Un requerimiento esencial de una medición de actitudes es que dichas afirmaciones estén en escala, esto es, que las afirmaciones sean seleccionadas y agrupadas de un número mucho mayor de afirmaciones de acuerdo con ciertas técnicas. Estas técnicas serán discutidas en el capítulo 18.

En encuestas, las preguntas sobre opiniones y actitudes presentan más problemas en su construcción que las preguntas fácticas. Es relativamente simple, por ejemplo, obtener información certera acerca de si una persona esta casada o soltera. Razonablemente, uno puede asumir que el encuestado sabe si esta casado o no. Cuando tratamos con opiniones o actitudes la suposición de si alguien sabe no siempre puede hacerse. Los encuestados pueden no tener una actitud con respecto a la legalidad del aborto, o si la tienen, puede estar latente. Lo que es más, dado que muchas actitudes tienen numerosos aspectos o dimensiones, los encuestados pueden estar de acuerdo con un aspecto, y no con otro. Es por esto que las actitudes no pueden ser medidas mediante una sola pregunta. Por ejemplo, si una persona esta fuertemente en desacuerdo con el enunciado "El aborto debería estar disponible para cualquier mujer que lo requiera", esto no implica una amplia actitud anti aborto.

La opinión de esta persona podría ser diferente si la vida de la mujer en cuestión estuviese en peligro, si el embarazo fuera el resultado del incesto o la violación, o si un doctor dijera que el bebe resultará severamente deformado. Mediante la utilización de varias afirmaciones, se pueden reducir los efectos de las respuestas de una sola dimensión.

Finalmente, respuestas a preguntas de opinión o actitud son más sensibles a los cambios en la formulación, el énfasis y la secuencia que las respuestas a preguntas fácticas. Esto refleja, en parte, la multidimensionalidad de muchas actitudes. Las preguntas presentadas en diferentes formas a veces reflejan diferentes aspectos de la actitud y así resultan en diversas respuestas.

Tipos de Preguntas

El contenido de la pregunta es solo un aspecto importante en la construcción de cuestionarios de encuesta. El investigador debe también considerar la estructura de la pregunta y el formato de las categorías de respuesta que acompañan la pregunta. Se puede distinguir entre tres tipos de estructura de pregunta: preguntas abiertas, preguntas cerradas y preguntas de contingencia.

Preguntas abiertas y preguntas cerradas

Las preguntas en un cuestionario pueden ser tanto abiertas como cerradas. En una **pregunta cerrada**, a los encuestados se les ofrece una serie de respuestas de las cuales deben elegir aquella que represente más fielmente sus puntos de vista. Por ejemplo, para medir el grado de satisfacción con la vida familiar, en la Encuesta Social

General, un sondeo realizado anualmente por el Consejo Nacional de Estudios de Opinión, fueron utilizadas las siguientes preguntas cerradas:

Señale el número que muestre el grado de satisfacción que obtiene de su vida familiar

1. Muchísima
2. Mucha
3. Bastante
4. La suficiente
5. Un poco
6. Muy poco
7. Nada
8. No sabe
9. No responde

Las respuestas a preguntas cerradas pueden ser también más elaboradas, como la siguiente pregunta, sacada de un sondeo acerca de las actitudes de hombres y mujeres acerca de el lugar y el rol de la mujer.³⁴

Suponga que ambos, un esposo y una esposa tienen trabajos buenos e interesantes y al esposo le ofrecen un muy buen trabajo en otra ciudad. Asumiendo que no tienen hijos, ¿cuál de estas soluciones piensa usted que deberían considerar?

- El esposo debería rechazar la oferta de trabajo
- La esposa debería renunciar y trasladarse con su esposo
- El esposo debería aceptar el nuevo trabajo y mudarse / la esposa debería mantener su trabajo y quedarse.

Las preguntas cerradas son fáciles de preguntar, y rápidas de responder; no requieren escritura de parte del encuestado, ni de parte del entrevistador, y su análisis es sencillo. Su mayor desventaja es que pueden introducir sesgos, ya sea forzando al encuestado a elegir entre alternativas dadas, o haciendo que el entrevistado seleccione entre alternativas que de otra manera podrían no haberle venido a la mente.

Las **preguntas abiertas** no son seguidas de ningún tipo de opción específica, y las respuestas de los entrevistados son grabadas por completo. Por ejemplo, la pregunta "Personalmente, ¿cuáles piensa usted que son los problemas más graves de los cuales el gobierno en Washington debería intentar resolver?" es una pregunta abierta utilizada frecuentemente en cuestionarios diseñados para estudiar la opinión pública. La ventaja de la pregunta abierta es que no obliga al encuestado a adaptarse a respuestas preconcebidas: habiendo entendido la intención de la pregunta, uno puede expresar sus pensamientos libremente, espontáneamente, y en su propio lenguaje. Si las respuestas a preguntas abiertas no son claras, el entrevistador puede profundizar pidiendo al encuestado que explique más o que de fundamentos para algo indicado anteriormente; tales preguntas abiertas permiten al entrevistador aclarar malos entendidos, y fomentan el *rapport*. Sin embargo, las preguntas abiertas son difíciles de responder y son más difíciles de analizar. El investigador debe diseñar un marco de códigos para poder clasificar las variadas respuestas; en este proceso, los detalles de la información provista por el encuestado puede perderse (ver capítulo 14).

La elección apropiada entre una pregunta abierta y una cerrada depende de un número de factores. Hace algunos años, Paul Lazarsfeld sugirió el uso de las siguientes consideraciones para determinar esta cuestión³⁵:

³⁴ Rita J. Simon Y Jean M. Landis, "Report: Woman's and men's attitudes about a woman's place and role", *Public Opinion Quarterly*, 53 (1989): 265-276.

³⁵ Paul Lazarsfeld, "The Controversy over Detailed Interviews: An Offer for Negotiation", *Public Opinion Quarterly*, 8 (1944): 38-60.

1. *Los objetivos del cuestionario.* Las preguntas cerradas son apropiadas cuando el objetivo del investigador es llevar al encuestado a expresar acuerdo o desacuerdo con un punto de vista explícito. Cuando el investigador desea conocer el proceso mediante el cual el encuestado llegó a un punto de vista en particular, una pregunta abierta sería más apropiada.

2. *El nivel de información del encuestado sobre el tema en cuestión.* Las preguntas abiertas proveen oportunidades para que el investigador pueda constatar la falta de información por parte del encuestado, mientras que no sucede lo mismo con las preguntas cerradas. Obviamente, es inútil realizar preguntas que están por encima de las experiencias de los encuestados.

3. *La medida en la que el tema ha sido reflexionado por el encuestado.* La pregunta abierta es preferible en situaciones en donde los encuestados aún no han cristalizado sus opiniones. La utilización de una pregunta cerrada en tales situaciones involucra el riesgo que al aceptar una de las alternativas ofrecidas, el encuestado pueda hacer una elección que sea bastante diferente a una opinión que hubieran expresado de haber pasado por el proceso de recordar y evaluar sus experiencias pasadas.

4. *La facilidad con la que el contenido de la respuesta pueda ser comunicado por el encuestado o la medida en que el encuestado es motivado a comunicar acerca del tema.* La pregunta cerrada requiere menos motivación para comunicar por parte del entrevistado, y la respuesta en sí misma es usualmente menos reveladora (y por lo tanto, menos amenazadora) que en el caso de una pregunta abierta. El investigador que utiliza preguntas cerradas tiende a encontrar menos negativas a responder.

A veces pueden haber buenas razones para realizar la misma pregunta tanto en forma abierta como cerrada. Por ejemplo, una respuesta abierta a la pregunta “¿Quién dirige América?” proveerá una idea clara de la percepción del sistema político del encuestado, y la significancia que la persona adjudica a los distintos grupos de poder. Aunque este dato es muy valorado, podría impedir las comparaciones entre las respuestas de un grupo de encuestados con otro. Lo que es más, uno no puede estar seguro que toda la información de importancia para el entrevistado haya sido mencionada; factores como la inhabilidad para articular ideas o un lapso momentario en la memoria pueden causar la omisión de puntos significativos. Por lo tanto, en el investigador puede volver a formular la misma pregunta otra vez, más adelante en la entrevista, pero esta vez en forma cerrada.

Preguntas de contingencia

Frecuentemente, las preguntas que son relevantes para un encuestado pueden ser irrelevantes para otro. Por ejemplo, la pregunta “Detalle las razones más importantes por las cuales planea ir a la universidad” obviamente es pertinente únicamente para estudiantes de escuela secundaria que están planeando ir a la universidad. Es frecuentemente necesario incluir preguntas que puedan aplicarse solamente a algunos encuestados y no a otros. Algunas preguntas pueden ser relevantes solo para mujeres y no para hombres, otras serán aplicadas solamente a encuestados que trabajan por cuenta propia, etc.

La **preguntas de contingencia** –un caso especial de pregunta cerrada- es aquella que solo se aplica a un subgrupo de encuestados. La relevancia de la pregunta para este subgrupo es determinada por la respuesta de todos los

encuestados a una **pregunta filtro** precedente. Por ejemplo, en un sondeo de los medios de noticias, la pregunta filtro podría ser, “¿Usted sigue las noticias en los diarios regularmente?”. La pregunta de contingencia podría ser “¿Sobre que evento reciente recuerda haber leído? (de una breve descripción)”. La relevancia de la segunda pregunta para el encuestado es contingente a su respuesta a la pregunta filtro. Solo los encuestados que hayan respondido “Sí” a la pregunta filtro encontrarán la pregunta de contingencia relevante. Por lo tanto, las categorías de respuesta a la pregunta filtro serán “1. Sí (responda la siguiente pregunta); 2. No (Saltee hasta la pregunta 3)”. Los formatos para las preguntas filtro y contingentes pueden variar. Una alternativa es escribir direcciones al lado de cada categoría de respuesta en la pregunta filtro. Otro formato común puede ser usar flechas para dirigir al encuestado a saltar hacia otra pregunta o responder la pregunta de contingencia, como en el ejemplo que sigue:

¿Es este el primer trabajo de tiempo completo que ha tenido desde su graduación de la universidad?

1. Sí
2. No



¿Qué sucedió con el trabajo que tenía antes? – fue promovido, despedido, qué sucedió? (Marque una opción)

1. La compañía cerró
2. Fue despedido
3. El trabajo se detuvo; era estacional
4. Renunció en forma voluntaria
5. Promocionado; trasladado
6. Otras.

Otro formato es encerrar la pregunta de contingencia y separarla de las preguntas ordinarias, que deben ser respondidas por todos. Un ejemplo de tal formato aparece en la Figura 11.1. Cuando hay varios subgrupos a los que el cuestionario está dirigido y cuando varias preguntas de contingencia pueden aplicarse a cada subgrupo, es útil indicar por número cuales preguntas el encuestado debería responder. Las instrucciones son escritas al lado de la categoría de respuesta apropiada en la pregunta filtro. Esto es demostrado en el siguiente ejemplo:

22. ¿Está usted buscando otro empleo en este momento?

- Sí
 - No
 - No sabe
 - Inapropiado
- } Ir a la pregunta 25.

Figura 11.1
Pregunta de contingencia

RESPONDA A LAS PREGUNTAS QUE SIGUEN SI USTED ES UN ESTUDIANTE DE ÚLTIMO AÑO DE LA ESCUELA SECUNDARIA PLANEANDO ASISTIR A LA UNIVERSIDAD EL AÑO PRÓXIMO. SINO, SALTEE A LA PREGUNTA 144.

137. ¿Rindió usted los exámenes de entrada a la universidad?

- Sí
- No

138. ¿Sabe usted en forma definitiva a que universidad asistirá?

- Sí
- No

139. Si su respuesta es "sí", ¿cómo se compara esta institución a otras que usted estaba considerando, en las siguientes formas?

1. 2. 3. 4. ofreciendo el curso de estudio que usted quiere

1. 2. 3. 4. la reputación general de la institución

SALTEE A LA PREGUNTA 151 EN LA SIGUIENTE PÁGINA

Formatos de pregunta

Ahora discutiremos algunas de las técnicas más comunes para estructurar las categorías de respuesta para las preguntas cerradas. El formato general es presentar todas las posibles respuestas, y que el encuestado marque las categorías apropiadas. El encuestado puede circular, escribir el número de la respuesta, o chequear un casillero o espacio en blanco, por ejemplo:

¿Cuál es su estado civil?

- | | | |
|--------------|-------------------------------------|---------------|
| - Casado | <input type="checkbox"/> Casado | 1. Casado |
| - Soltero | <input type="checkbox"/> Soltero | 2. Soltero |
| - Divorciado | <input type="checkbox"/> Divorciado | 3. Divorciado |
| - Viudo | <input type="checkbox"/> Viudo | 4. Viudo |

Por supuesto, las instrucciones específicas deberían ser provistas. Entre los tres métodos mostrados, el menos recomendable es el que utiliza espacios en blanco, ya que los encuestados podrían chequear entre los espacios, haciendo difícil entender cual categoría se tuvo la intención de marcar. Circular un número de código es preferible, ya que es fácil de transferir a la computadora.

Clasificación

Uno de los formatos más comunes para preguntas hechas en encuestas de ciencias sociales es la escala de clasificación. La **escala de clasificación** es utilizada cuando los encuestados deben realizar un juicio en relación con grupos de categorías ordenadas, tales como "fuertemente de acuerdo", "favorable" o "muy a menudo"; por ejemplo:

La policía debería tener el poder de realizar una búsqueda completa de cualquier motorista arrestado por una ofensa tal como exceso de velocidad.

1. Fuertemente de acuerdo
2. De acuerdo
3. En desacuerdo
4. Fuertemente en desacuerdo
5. Sin opinión

Las categorías de respuesta de tales preguntas se denominan **cuantificadores** ; reflejan la intensidad de un juicio particular involucrado. Los siguientes grupos de categorías de respuesta son bastante comunes:

- | | | |
|------------------------------|-------------------|----------|
| 1. Fuertemente de acuerdo | 1. Demasiado poco | 1. Más |
| 2. De acuerdo | 2. Correcto | 2. Igual |
| 3. Depende | 3. Demasiado | 3. Menos |
| 4. En desacuerdo | | |
| 5. Fuertemente en desacuerdo | | |

Los códigos numéricos que acompañan estas categorías son usualmente interpretados para representar la intensidad de la categoría de respuesta, de modo que a mayor número mayor intensidad en la respuesta. Sin embargo, deberíamos enfatizar que aunque asumimos que los cuantificadores involucrados están ordenados de acuerdo con su intensidad, esto no implica que la distancia entre ellos sea la misma. Efectivamente, las escalas de clasificación como estas suelen ser medidas mediante un nivel ordinal de medida, como es discutido en el capítulo 7.

A pesar de la dificultad para estimar intensidades, usualmente no podemos preguntar a los encuestados por estimados exactos porque a la mayoría le resultaría una tarea muy difícil. Mientras que parece relativamente fácil decir cuantas horas de televisión miró una persona en la semana, la mayoría de la gente tendrá una mayor dificultad estimando en forma precisa eventos de relativamente baja notabilidad, como las actitudes acerca de la política internacional.³⁶

Preguntas de matriz

La pregunta de matriz es un método para organizar un grupo grande de preguntas de clasificación que tienen las mismas categorías de respuesta. El siguiente es un ejemplo de tal mecanismo:

³⁶ Norman M. Bradburn y Seymour Sudman, *Improving Interview Method and Questionnaire Design* (San Francisco: Jossey-Bass, 1974), pp. 152-162.

Indique su reacción a las siguientes afirmaciones

	Fuertemente de acuerdo	De acuerdo	Depende	En desacuerdo	Fuertemente en desacuerdo
Mi voto me otorga todo el poder que quiero con respecto a asuntos gubernamentales.	()	()	()	()	()
Si me quejara con la gente en la agencia de la ciudad ellos arreglarían lo que sea que esté mal.	()	()	()	()	()
A veces he deseado que los oficiales del gobierno prestaran más atención a lo que pienso.	()	()	()	()	()

Ordenamiento de cartas

Otra forma de medir intensidades en los juicios es mediante el **ordenamiento de cartas**. Al entrevistado se le entregan una serie de tarjetas, cada cual con un enunciado, y luego se le pide que las ubique en una de siete casillas, dependiendo en su grado de acuerdo o desacuerdo con el enunciado. El que sigue es un ejemplo de tal mecanismo:

Cuan fuertemente está usted de acuerdo o en desacuerdo con el siguiente enunciado: La gente que es capaz de trabajar pero no lo hacen son un peso para la sociedad.						
Fuertemente de acuerdo						Fuertemente en desacuerdo
+++	++	+	+ -	-	--	---

Diferencial Semántico

El diferencial semántico es otro tipo de escala de clasificación. Mide la reacción del encuestado a cierto objeto o concepto, en términos de una clasificación en escalas bipolares definidas por adjetivos contrastantes en cada extremo³⁷:

Bien 3 2 1 0 1 2 3 Mal

³⁷ David R. Heise, "The Semantic Differential and Attitude Research", en *Attitude Measurement*, ed. Gene F. Summers (Stokie: Ill.: Rand McNally, 1970), p. 235.

El número cero marca la posición neutral en la escala, y las posiciones 1-3 miden la intensidad en ambas direcciones, en donde el 1 es la reacción menor, y el 3 es la reacción más intensa.

Examine este ejemplo de la aplicación del diferencial semántico.³⁸

Aquí hay una lista de palabras utilizadas para describir funcionarios civiles. Entre cada par, hay línea de medidas de siete columnas. Tomando el primer par de palabras - por ejemplo, "bueno/malo"- como ejemplo, columna en la extrema izquierda significaría que el funcionario civil es muy bueno, la línea de al lado significaría que es bastante bueno, y así sucesivamente. Las palabras en la parte superior de su tarjeta lo ayudarán a elegir la columna que usted piense que es la más apropiada.

Ahora, dígame, ¿cuál columna usted utilizaría para describir a los funcionarios civiles?

	Muy	Bastante	Un poco	Ni	Un poco	Bastante	Muy	
Bueno	—:	—:	—:	—:	—:	—:	—:	Malo
Honesto	—:	—:	—:	—:	—:	—:	—:	Deshonesto
Eficiente	—:	—:	—:	—:	—:	—:	—:	Ineficiente
Profundo	—:	—:	—:	—:	—:	—:	—:	Superficial
Activo	—:	—:	—:	—:	—:	—:	—:	Pasivo

Ordenamiento

Se utiliza el ordenamiento en cuestionarios cuando buscamos obtener información con respecto al grado de importancia de las prioridades que la gente atribuye a un grupo de actitudes u objetos. Por ejemplo, en una estadística acerca de la calidad de vida, se les pidió a los encuestados que ordenaran varias dimensiones que consideraran importantes en la vida:

“Quisiera que me dijera que es lo que usted encuentra como importante en la vida. Por favor, mire esta tarjeta y dígame cual de estos enunciados es más importante para *usted* como meta para *su* vida, cuál lo sigue en importancia, cuál está en tercer lugar y cual en cuarto.”

	Rango			
	1	2	3	4
Una vida prospera (tener un buen ingreso y ser capaz de disfrutar de las buenas cosas de la vida)				
Una vida familiar (una vida completamente centrada en mi familia)				
Una vida importante (una vida de logros que me brinden respeto y reconocimiento)				
Una vida segura (asegurarme de que todas mis necesidades básicas y gastos estén provistos)				

El ordenamiento es un mecanismo muy útil para dar algún sentido de orden relativo entre objetos o juicios. Esto es particularmente importante dado que a muchas de las propiedades medidas en ciencias sociales (por ejemplo, “calidad de vida”, “estatus”) no se les puede otorgar un valor numérico preciso. No obstante, con el uso del ordenamiento podemos al menos obtener información con respecto a su orden relativo. Debemos enfatizar, sin embargo, que el ordenamiento no provee información

³⁸ David Nachmias y David H. Rosenbloom, *Bureaucratic Culture: Citizens and Administrators in Israel*, (New York: St. Martin's Press, 1978), pp. 110-115.

acerca de la distancia entre los rangos. La diferencia entre, por ejemplo, el rango 1 y el rango 2 puede no ser la misma que la diferencia entre el rango 2 y el rango 3.

Secuencia de las preguntas

Luego que se ha determinado el formato de las preguntas, se debe considerar el orden en el que las mismas son presentadas en el cuestionario. Existen dos patrones generales de secuencias que se consideran más apropiados para motivar a los entrevistados a cooperar: la secuencia de embudo y la secuencia de embudo invertido.

Secuencia de embudo

En la secuencia de embudo, cada pregunta sucesiva está en relación con la pregunta previa, y tiene un alcance progresivamente más estrecho. Por ejemplo, si uno estuviera interesado en averiguar como las visiones de los encuestados sobre problemas políticos, económicos y sociales están relacionadas con los periódicos que leen, uno podría querer saber que es lo que los encuestados consideran problemático, cual es su percepción sobre la relativa importancia de cada problema, cuanta información tienen sobre el tema, cuales son sus recursos de información, y si ciertos periódicos han influenciado su pensamiento acerca del problema. Las siguientes preguntas forman una secuencia de embudo:

1. ¿Cuáles cree usted que son los principales problemas a los que se enfrenta la Nación?
2. De todos los problemas que acaba de mencionar, ¿cuál cree usted que es el más importante?
3. ¿De donde ha sacado la mayor parte de la información que usted tiene sobre este problema?
4. ¿Lee usted el *Washington Post*?

Cuando el objetivo de la encuesta es obtener información detallada, y cuando el encuestado es motivado a proveer tal información, la secuencia de embudo ayuda a que el encuestado recuerde detalles más eficientemente. Lo que es más, formulando la pregunta más amplia en primer lugar, el encuestado puede evitar imponer un marco de referencia antes de obtener la perspectiva del encuestado. Cuando el objetivo de una encuesta es descubrir respuestas no anticipadas, las preguntas más amplias deberían formularse en primer lugar.³⁹

Secuencia de embudo invertido

En la secuencia del embudo invertido, las preguntas más puntuales son seguidas de preguntas más amplias. Cuando el tema de la encuesta no motiva fuertemente a los encuestados brindar información –ya sea porque el tema no es importante para ellos o porque sus experiencias no son lo suficientemente recientes para estar vividas en sus memorias- puede ser de ayuda comenzar con preguntas más puntuales, que son más fáciles de responder, y reservar las preguntas más amplias (y más difíciles) para más tarde. Si el propósito es obtener una generalización en la forma de un juicio con respecto a una situación concreta y si el encuestador no está familiarizado con los hechos pero el encuestado los conoce, las preguntas más

³⁹ Raymod L. Gorden, *Interviewing: Strategy, Techniques, and Tactics*, 3^o ed. (Homewood, Ill.: Dorsey, 1980), pp. 415-416.

puntuales, destinadas a establecer datos específicos, deben preceder a las preguntas que requieran un juicio generalizado.⁴⁰

En el siguiente ejemplo, se intenta obtener el juicio de los encuestados con respecto a la efectividad de las operaciones de rescate durante un desastre. Para ayudar a la gente a realizar un juicio imparcial, el investigador sintió que era conveniente tratar con temas específicos en primer lugar, preguntando sobre cuestiones generales después.⁴¹

1. ¿Cuánta gente murió en el tornado?
2. ¿Cuánta gente supone usted que fue herida tan seriamente que necesitaron ir al hospital?
3. ¿Cuánto tiempo pasó hasta que la gente que lo necesitaba llegó al hospital?
4. ¿Vio usted a alguien administrar primeros auxilios, dando respiración artificial o deteniendo hemorragias?
5. En general, ¿cuán eficientemente cree usted que las operaciones de primeros auxilios y de rescate fueron llevadas a cabo?

Se ha demostrado que el orden en que las preguntas son presentadas afecta el tipo de respuesta dada. Por ejemplo, hay evidencia de que las respuestas a preguntas de actitud en encuestas pueden variar marcadamente, dependiendo de los ítems precedentes en el cuestionario. En un estudio reciente, se preguntó a más de 1100 encuestados acerca de cuestiones tales como el aborto, el gasto en defensa, y las prestaciones sociales.⁴² En una versión del cuestionario, estas preguntas fueron precedidas por preguntas relacionadas con el contexto; en otras, las preguntas fueron precedidas por preguntas neutrales. Por ejemplo, la pregunta sobre el aborto “Usted está a favor o en contra de la decisión de la Corte Suprema de legalizar el aborto?” fue precedida, en la primer versión, por un número de preguntas contextuales acerca de los valores tradicionales y la violación. Los encuestados fueron generalmente afectados por las preguntas contextuales relacionadas, especialmente cuando tenían opiniones encontradas acerca del tema principal. Hay también evidencia de que la posición de un ítem en una lista tiene un impacto significativo sobre la elección o no del mismo. Los ítems que aparecen en primer lugar suelen ser elegidos más a menudo.⁴³ Ha sido demostrado también que cuando se pide que los encuestados asignen un valor numérico a un grupo de ítems (por ejemplo, de acuerdo a su grado de importancia), los ítems que aparecen primero suelen recibir un rango mayor.

En la siguiente pregunta, es más probable que los encuestados asignen el primer rango a la primer categoría que a la última simplemente porque está listada en primer lugar.

Entre los ítems que aparecen abajo, ¿qué se requiere para ser importante y admirado por otros estudiantes aquí en la escuela? (Rango del 1 al 6).

- Provenir de la familia correcta
- Ser un líder en las actividades
- Tener un buen auto
- Tener notas altas
- Ser una estrella atlética
- Ser popular

⁴⁰ Ibid.

⁴¹ Ibid.

⁴² Roger Tourangeau, Kenneth A. Rasinski, Norman M. Bradburn y Roy D'Andrade, “Carryover Effects in Attitude Surveys”, *Public Opinion Quarterly*, 53 (1989): 495-524.

⁴³ William A. Belson, “The Effects of Reversing the Presentation Order on Verbal Rating Scales”, *Journal of Advertising Research*, 6:4 (1966): 30-37.

Este problema puede surgir especialmente en situaciones donde las preguntas son enunciados subjetivos como actitudes que no son centrales para el encuestado. En tales situaciones, el ítem que aparece primero tiene a formar un punto de referencia para todos los ítems que siguen. Este problema puede resolverse familiarizando a los encuestados con la lista de ítems antes de realizar las evaluaciones. Alternativamente, el orden de la presentación puede ser aleatorio, para que los efectos del orden resulten aleatorios también, y de esta forma no resulte en un sesgo sistematizado.⁴⁴

Finalmente, debería ser señalado que las preguntas que son presentadas primero en el cuestionario deberían poner a gusto al encuestado; aquellos en una entrevista deberían ayudar a crear *rapport* entre el encuestador y el encuestado. Así, la pregunta que abra debería ser fácil de responder, interesante y no controversial. Por ejemplo, preguntas sobre los hábitos de bebida de un entrevistado o sobre su vida sexual, si son ubicadas en primer lugar, probablemente eleven la tasa de rechazo. Es también recomendado que las preguntas abiertas sean ubicadas al final, ya que usualmente requieren más tiempo y reflexión y podrían reducir la motivación inicial del encuestado para cooperar.

Evitando sesgos: Trampas en la construcción de cuestionarios

Selección de palabras

La pregunta debe ser formulada de modo tal que el encuestado pueda entenderla. Por ejemplo, el vocabulario del investigador podría incluir palabras tales como *carismático* que podrían no ser entendidas por la mayoría de las personas. Si los encuestados vienen de diversas procedencias, el investigador debería utilizar palabras que puedan ser entendidas por un estudiante promedio del octavo grado. Lo que es más, palabras que pueden estar sujetas a interpretación deberían ser evitadas o aclaradas. Por ejemplo, preguntar si alguien es liberal podría referir a la educación de la persona, a sus opiniones políticas, profesión o vida sexual. Pero una pregunta como “¿Usted se considera liberal, políticamente hablando?” instruye al encuestado a utilizar un marco de referencia político para responder la pregunta. Las preguntas deberían ser formuladas de modo que el encuestado comprenda la pregunta y de modo tal que tenga el mismo significado para todas las personas que deban responderla.

Respuestas en masa

La respuesta en masa es la tendencia a responder todas las preguntas en una dirección específica sin importar su contenido.⁴⁵ Esto puede resultar problemático cuando un grupo de preguntas son presentadas en conjunto con el mismo tipo de formato de respuesta, especialmente las preguntas refieren todas al mismo tema. Por ejemplo, si un grupo de preguntas reflejan una actitud en contra del aborto, los encuestados que comparten esta opinión podrían elegir todas las mismas categorías de respuesta (por ejemplo, todos “Fuertemente en desacuerdo” o “fuertemente de acuerdo”) simplemente porque asumen que estas categorías expresarán objeciones al aborto. La respuesta en masa puede ser evitada cambiando el formato de la pregunta,

⁴⁴ Edwin H. Carpenter y Larry G. Blackwood, “The Effects of Question Position on Responses to Attitudinal Questions”, *Rural Sociology*, 44 (1979): 56-72.

⁴⁵ Kenneth D. Bailey, *Methods of Social Research* (Nueva York: Free Press, 1987), p. 133.

variando las categorías de respuesta para cada pregunta o evitando juntar a las preguntas que refieran a un mismo tema.

Preguntas tendenciosas

Una **pregunta tendenciosa** es una pregunta formulada de tal forma que sugiere al encuestado que el investigador espera una respuesta determinada. Una pregunta diseñada para obtener actitudes generales con respecto al aborto podría ser “Usted está a favor o en contra del aborto legal?”. La misma pregunta formulada en forma tendenciosa podría ser, “¿Usted no diría que está a favor del aborto legal, no?. O de un modo más sutil, una pregunta tendenciosa podría ser, “¿diría usted que está en contra del aborto legal?. Esta última pregunta hace más fácil que los encuestados respondan sí en lugar de no ya que la mayoría de la gente se siente más cómoda acordando con el lenguaje de la pregunta y no contradiciendo al encuestador.

Los encuestados también tienden a estar de acuerdo con preguntas que sostienen las normas aceptadas o que son percibidas como socialmente deseables. Las preguntas que reflejan comportamientos o actitudes no deseables socialmente son elegidas menos frecuentemente que aquellas que se encuentran más arriba en la escala de aceptabilidad social. De forma similar, el etiquetamiento de problemas y mejoras pueden tener un efecto sustancial en el apoyo público de algunos temas. Los análisis de variaciones en las palabras de los cuestionarios en la Encuesta Social General mostró diferencias significativas en las respuestas. Por ejemplo, cuando una pregunta sobre el gasto en asistencia social decía “¿estamos gastando demasiado, demasiado poco, o la cantidad justa en asistencia social?” 23 por ciento dijo demasiado poco. Sin embargo, cuando la pregunta fue formulada como “¿estamos gastando demasiado poco, demasiado o la cantidad justa en la asistencia a los pobres?” casi el 63 por ciento dijo demasiado poco.⁴⁶

Las preguntas tendenciosas deben ser evitadas si uno está buscando respuestas no distorsionadas. Bajo ciertas circunstancias, sin embargo, las preguntas tendenciosas pueden servir al objetivo de la investigación. La pregunta “¿Usted estaría a favor de mandar comida al extranjero, para alimentar a la gente hambrienta en la India?” fue utilizada para determinar el número de gente que estaba tan fuertemente en desacuerdo a enviar alimento a otros países que rechazaban la idea aún bajo el contexto fuertemente emocional de “gente hambrienta”.⁴⁷

Preguntas amenazadoras

A menudo es necesario incluir preguntas o temas que el encuestado puede encontrar vergonzosas y por lo tanto difíciles de responder. Tales **preguntas amenazadoras** son, de acuerdo con Norman Bradburn y sus coautores, “preguntas que despiertan ansiedad acerca de, por ejemplo, comportamientos que son ilegales o contra-normativos o sobre comportamientos que, aunque no socialmente desviados, no son usualmente discutidos en público sin cierta tensión.”⁴⁸ Las preguntas amenazadoras podrían preguntar, por ejemplo, acerca de los hábitos de juego, bebida o preferencias sexuales.

Hay considerable evidencia empírica de que las preguntas amenazantes llevan a sesgos en las respuestas, ya sea la negación del comportamiento en cuestión o quitándole importancia. En general, el reporte de ciertos comportamientos decrece a medida que las preguntas aumentan el grado de amenaza. Al ser presentados con una

⁴⁶ Kenneth A. Rasinski, “The Effect of Question Wording on Public Support for Government Spending”, *Public Opinion Quarterly*, 53 (1989): 388-394.

⁴⁷ Robert I. Kahn y Charles F. Cannell, *The Dynamics of Interviewing*, (Nueva York: Wiley, 1957), p. 129.

⁴⁸ Norman M. Bradburn, Seymour Sudman, Ed Blair y Carol Stocking, “Question Threat and Response Bias”, *Public Opinion Quarterly*, 42 (1978): 221-222.

pregunta amenazante, los encuestados se ven atrapados en un conflicto entre las demandas del rol del “buen encuestado”, quien responde sinceramente a todas las preguntas, y la tendencia a presentarse a si mismo en forma positiva. El conflicto es usualmente resuelto no mediante la negativa a responder, sino reportando que uno no se involucró en una actividad en particular, cuando uno de hecho si lo hizo.⁴⁹

Ya que las preguntas amenazantes pueden causar respuestas sesgadas, es importante que los investigadores identifiquen primero si ciertas preguntas son amenazadoras o no. Norman Bradburn y Seymour Sudman sugieren que el mejor método para determinar la amenaza relativa de las preguntas es pidiendo a los encuestados que evalúen los tópicos de las preguntas de acuerdo a cuan incómodas creen que la mayoría de las personas se sentirían al hablar de ellos.⁵⁰ Uno podría también preguntar acerca de las reacciones de los propios encuestados a las preguntas, o evaluar el grado de dificultad que los tópicos causaron en la encuesta.

Una vez que las preguntas amenazadoras han sido identificadas, ¿qué se debería hacer con ellas? En un estudio exhaustivo que trató acerca de los efectos de respuesta a preguntas amenazadoras en las investigaciones por encuesta, Bradburn y Sudman determinaron que la construcción de las preguntas hace una gran diferencia.⁵¹ Quizás su descubrimiento más significativo haya sido el descubrimiento de que la exactitud de las respuestas aumenta en forma significativa al utilizar una larga introducción a la pregunta en lugar de realizar preguntas cortas, utilizando preguntas abiertas en lugar de preguntas cerradas y, en menor medida, dejando que el entrevistado elija sus propias palabras al tratar temas que pudieran resultar sensibles. Su cuestionario contenía un ítem acerca de la cantidad de veces que el encuestado había estado intoxicado durante el último año. En su versión corta y cerrada, la pregunta decía: “En el pasado año, ¿cuán a menudo se intoxicó usted mediante la consumición de bebidas alcohólicas de cualquier tipo?. Los encuestados debían luego clasificar su respuesta en una de las siguientes categorías: nunca, una vez al año o menos, cada algunos meses, una vez al mes, cada algunas semanas, una vez por semana, varias veces por semana y a diario. En la versión más larga y abierta, primero se pedía a los encuestados que definieran intoxicación en sus propias palabras: “A veces la gente bebe demasiada cerveza, vino o whisky y como resultado actúan en forma diferente a como actúan normalmente. Qué palabra cree usted que se debería utilizar para describir a este tipo de gente, de modo que usted entienda lo que queremos decir y se sienta cómodo hablando al respecto?. Luego, la pregunta sobre la intoxicación leía: “Ocasionalmente la gente bebe con el estómago vacío o bebe demasiado y termina (utilizar la palabra del encuestado). En el pasado año, ¿cuán menudo a resultado usted (palabra del encuestado) al beber cualquier tipo de bebida alcohólica?” No se dieron categorías de repuesta para esta pregunta.⁵²

Preguntas de múltiples opciones

Las preguntas de múltiples opciones combinan dos o más preguntas en una. Aquí presentamos un ejemplo sacado de un sondeo de opinión sobre la violencia doméstica:

⁴⁹ Ibid., pp. 221-234.

⁵⁰ Bradburn y Sudman, *Improving Interview Method and Questionnaire Design*, p. 165.

⁵¹ Ibid. pp. 14-25.

⁵² Ibid., p. 18.

La violencia doméstica y el SIDA son los problemas más serios a los que se enfrenta América hoy en día:

- De acuerdo
- Depende
- En desacuerdo
- Fuertemente en desacuerdo

El problema con una pregunta de este tipo es que podría confundir a los encuestados que estén de acuerdo con una parte de la pregunta, por ejemplo, la violencia doméstica, pero estén en desacuerdo con la otra, SIDA. Muchas preguntas que contienen la palabra y son posiblemente de opción múltiple. Este tipo de preguntas pueden ser utilizadas, sin embargo, si las dimensiones separadas por el y son mutuamente excluyentes y al encuestado se le pide que seleccione una opción, o las ordene de acuerdo con algún tipo de criterio, por ejemplo:

En el momento presente, el país está enfrentando dos graves problemas: el medio ambiente y la violencia doméstica. ¿Cuál de estos problemas diría usted que es el más importante?

- El medio ambiente
- La violencia doméstica

Carta de presentación

Una vez que el cuestionario ha sido construido, el paso siguiente es escribir un enunciado introductorio (para una encuesta telefónica) o una carta de presentación (para un cuestionario por correo) para explicar el propósito de la encuesta a los encuestados y para fomentar una alta tasa de respuestas. Esto es particularmente importante en los cuestionarios por correo, donde la dificultad de asegurar una alta tasa de respuestas, especialmente cuando se necesita realizar más que algunas preguntas simples, es bien documentado (ver capítulo 10).

Una carta de presentación debe poder vencer cualquier tipo de resistencia o prejuicio que el encuestado pueda tener con respecto a la encuesta. Debería (1) identificar la organización patrocinadora o las personas realizando el estudio, (2) explicar el propósito del estudio, (3) explicar el porque es tan importante que el encuestado responda el cuestionario, y (4) asegurar al encuestado que la información provista será mantenida en estricta confidencia.

En general, la carta de presentación para un cuestionario por correo debe ser más detallada que el enunciado introductorio en una entrevista personal. En una encuesta, el encuestador siempre se encuentra presente para explicar o persuadir al encuestado, en caso de ser necesario. Con un cuestionario por correo, la carta de presentación es todo lo que hay, y por lo tanto, su función es de extrema importancia.

Dos ejemplos de cartas de presentación utilizadas en varias encuestas por correo serán mostradas aquí. La primera, presentada en la Figura 11.2 fue utilizada en un cuestionario por correo diseñado por el Instituto de Investigación Social en la Universidad Estatal de Florida bajo los auspicios del State Department of Manpower Planning of Florida para evaluar el acta de Empleados de Servicio Público y Entrenamiento, título VI (CETA).⁵³

El segundo ejemplo, reimpresso en la Figura 11.3 es de un estudio sobre el compromiso con las libertades civiles, conducido por investigadores en la universidad

⁵³ Mickey L. Burnim, *An Evaluation of the Public Service Employment Projects in Florida Created under Title VI of the Comprehensive Employment and Training Act of 1973* (Tallahasee: Universidad Estatal de Florida, 1978), p. 164.

de Wisconsin en Milwaukee.⁵⁴ La carta hace énfasis en la confidencialidad del estudio y explica en detalle como serán utilizadas las respuestas individuales.

Finalmente, un tema importante es el estilo utilizado en la carta de presentación, esto es, si es una carta formal o semi personal. En los dos ejemplos, un modelo de carta fue enviada a todos los encuestados incluidos en la muestra. Alternativamente, en lugar de enviar la carta a “Querido Amigo”, o “Querido Encuestado”, el nombre y domicilio del encuestado puede ser insertado. Ha sido demostrado que una carta más personal genera una tasa de respuesta ligeramente superior que una carta de muestra.⁵⁵

Figura 11.2

Carta de presentación Cuestionario Florida

A los operadores del programa:

La Office of Manpower Planning, Departamento de Asuntos Comunitarios, en conjunto con el State Manpower Services Council, ha fundado una evaluación especial del empleo de proyectos de servicios públicos autorizados bajo el Título VI del Comprehensive Employment and Training Act. Esta evaluación está siendo conducida por el Dr. M.L. Burnim en el Instituto de Investigación Social de la Universidad Estatal de Florida. El propósito de la evaluación es determinar el impacto del empleo de proyectos de servicios públicos en personas desempleadas en el estado de Florida y medir el beneficio de estos proyectos a las comunidades en las cuales son conducidos.

Como Ud. sabe, el empleo de servicio público es una parte central de la estrategia federal, estatal y local para superar los problemas de empleo e ingresos de aquellos los económicamente desaventajados y los desempleados. No hay duda de que el programa es necesitado en todo el país para crear oportunidades de empleo y entrenamiento para el gran número de personas que continúan desempleadas. Probablemente, Ud. esté al tanto, sin embargo, que los programas de empleo del servicio público son bastante controversiales y su futuro podría estar en peligro. Parte del motivo por el cual estos programas son tan controversiales es que no se ha conducido nunca una evaluación sistemática de los beneficios de estos programas para los individuos empleados y las comunidades involucradas.

Ya que esta evaluación específica tiene significativas implicancias para las políticas nacionales, insto fuertemente que Ud. asista al equipo de investigación para poder recopilar los datos necesarios. Es muy importante que Ud. complete el cuestionario que le será transmitido lo antes posible.

Gracias por su cooperación.

Sinceramente,

Edward A. Feaver, Director
Office of Manpower Planning

Figura 11.3

Carta de presentación Cuestionario de Wisconsin

Querido Amigo:

Estamos conduciendo un sondeo patrocinado por la Universidad de Wisconsin – Milwaukee y asistido por la Unión Americana de Libertades Civiles (UALC). Nuestro propósito es aprender más acerca de cómo la gente, como usted siente acerca de ciertos aspectos de las libertades civiles y como las creencias están relacionadas con el comportamiento. Usted ha sido seleccionado al azar para participar en este sondeo –así, sus opiniones representarán las

⁵⁴ Richard D. Bingham y James L. Gibson, “Conditions of Commitment to Civil Liberties”, sin publicación (Milwaukee: Department of Political Science, Universidad de Wisconsin, 1979).

⁵⁵ Michael T. Mattheson, “Type of Transmittal Letter and Questionnaire Color as Two Variables Influencing Response Rates in a Mail Survey”, *Journal of Applied Psychology*, 59 (1974): 532-536.

opiniones de miles de personas como usted.

Adjunta encontrará una copia de nuestro cuestionario. Si bien es un poco largo y requerirá unos 20 minutos para completarlo, esperamos que se pueda tomar el tiempo para completarlo y devolvernos el cuestionario en el sobre adjunto. La información que usted provea contribuirá a un importante estudio y podrá ser también utilizada para influenciar las políticas de la UALC.

Un poco acerca de la confidencialidad. Prometemos su confidencialidad bajo los estándares éticos académicos de la Asociación Americana de Ciencia Política. Su nombre no será revelado o asociado con su respuesta y nadie por fuera del staff del proyecto aquí en la Universidad de Wisconsin –Milwaukee será autorizado a ver sus respuestas. De este modo, si el UALC estuviera interesado en las implicancias políticas de nuestro estudio, no serán provistos de ningún tipo de información que lo identifique a usted como individuo. Por favor, note el número en la esquina superior derecha de su cuestionario. Este número nos permitirá identificarlo en forma provisoria. Refiriéndonos a este número sabremos que usted respondió este cuestionario y no tendremos que enviarle el correo de seguimiento que mandaremos a quienes no respondan.

Apreciamos su buena voluntad para ayudarnos en nuestros esfuerzos de investigación. Si usted quisiera una copia de nuestro estudio completo por favor indíquelo en la última página del cuestionario. Nos aseguraremos de que reciba una copia de nuestros resultados. Creemos que encontrará al cuestionario tanto interesante como provocativo y esperamos recibir su respuesta.

Sinceramente suyo,

Richard D. Bingham
Profesor asociado

James L. Gibson
Profesor asistente

Instrucciones

Otro elemento que debe ser considerado al construir un cuestionario son las instrucciones que van con cada pregunta o con un grupo de preguntas. Las instrucciones deberían ser incluidas siempre que las preguntas no se expliquen a sí mismas; pueden variar desde preguntas simples como “circule la categoría apropiada” a instrucciones más complejas que expliquen como ordenar un grupo de prioridades. Cuando el cuestionario es administrado por un encuestador, las instrucciones suelen estar escritas para el encuestador y por lo tanto suelen ser cortas y concisas, instruyendo al encuestador sobre que hacer cuando el encuestado provee cierta respuesta, cuando indagar sobre información más detallada, o como clarificar determinadas preguntas. El que sigue es un ejemplo de instrucciones escritas para el encuestador:

¿Quién fue su empleado en su último trabajo?
(INDAGAR SOBRE LA CATEGORÍA CORRECTA)

Privado
Ciudad
País
Estado
Federal
Trabajo por cuenta propia
Publico, sin fines de lucro
Otros _____ (especificar)
No sabe

Mientras que en un estudio de este tipo el encuestador esta disponible para responder preguntas que en encuestado pudiera tener, este no es el caso con cuestionarios por correo, donde cualquier pregunta que sea imprecisa o poco clara probablemente sea respondida incorrectamente, si es que es respondida. Por lo tanto, proveer instrucciones claras es extremadamente importante. Pueden variar desde instrucciones generales introduciendo el cuestionario o sus sub secciones hasta detalles especificos precediendo preguntas individuales.

El que sigue es un ejemplo de instrucciones generales dadas al principio de un cuestionario sobre actitudes acerca de las libertades civiles.⁵⁶

INSTRUCCIONES: Para cada una de las siguientes preguntas por favor marque la respuesta que más se acerque a sus sentimientos sobre el tema. No hay respuestas "correctas" o "incorrectas" –por favor responda las preguntas de la forma más honesta que le sea posible. Responda cada pregunta en el orden en el que aparece. Si desea hacer algún comentario adicional sobre cualquier pregunta específica o sobre los temas en general, utilice el espacio la final del cuestionario. Sus opiniones son extremadamente importantes para comprender estas complejas libertades civiles –¡estamos muy agradecidos de su cooperación!

El siguiente ejemplo, para el mismo cuestionario, introduce una sub sección, presentada en forma de matriz:

Como usted sabe, hay muchos grupos en América que intentan que el gobierno vea las cosas a su manera. Nos gustaría averiguar su opinión acerca cuales percibe usted que son sus propósitos, objetivos o ideas defendidas por estos grupos. En particular, nos gustaría saber su opinión sobre cuan significativo debería ser el cambio en el sistema de gobierno americano si las ideas de estos grupos se pusieran en práctica. Por favor califique cada uno de los siguientes grupos en términos de la naturaleza de los cambios en nuestro sistema de gobierno que se seguirían a la implementación de sus ideas.

Marque solo una respuesta en cada columna

	Comunistas	Nazis	Ku Klux Klan
<i>Las ideas, de ser implementadas, crearían un sistema de gobierno totalmente diferente y mucho peor.....</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<i>Las ideas, de ser implementadas, cambiarían significativamente nuestro sistema de gobierno para peor.....</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<i>Estoy en contra de las ideas, pero no cambiarían nuestro sistema de gobierno si fueran implementadas.....</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<i>Apoyo las ideas, pero ellas no cambiarían nuestro sistema de gobierno si fueran implementadas.....</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<i>Las ideas, de ser implementadas, cambiarían significativamente nuestro sistema de gobierno para mejor.....</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<i>Las ideas, si fueran implementadas, crearían un sistema de gobierno totalmente diferente y mucho mejor.....</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<i>Sin opinión.....</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Finalmente, aquí hay un ejemplo sobre una instrucción específica, para responder una única pregunta.

⁵⁶ Bingham y Gibson, "Conditions on Commitment to Civil Liberties".

¿En cuantos estados ha vivido usted en su vida? (cuente solo aquellos estados en los que haya residido por lo menos un año)

Construyendo un cuestionario: un caso de estudio

Hay muchas etapas involucradas en la construcción de un cuestionario, comenzando por el problema de investigación, hasta el proceso de formular las preguntas y considerar el formato y tipo de preguntas a utilizar. Para ilustrar esto presentamos en la Figura 11.4 un cuestionario basado en un estudio real, conducido por el Instituto de Investigación Social de la Universidad de Michigan.⁵⁷

El objetivo de este estudio era explorar las actitudes y percepciones relacionadas a problemas urbanos y relaciones raciales en 15 ciudades del norte de los Estados Unidos. Buscaba definir las características sociales y psicológicas así como las aspiraciones de las poblaciones urbanas blancas y negras. Fueron seleccionadas, en cada una de las ciudades en el estudio, una muestra blanca y otra negra. Se encuestó a aproximadamente 175 personas de raza negra y 175 personas de raza blanca en cada ciudad. Adicionalmente, 366 personas blancas fueron entrevistados en áreas suburbanas. En total, se entrevistó a 2809 personas negras y 2950 blancas. Los individuos entrevistados tenían edades entre los 16 y los 69 años y residían en viviendas privadas. El estudio utilizó dos tipos de cuestionarios, uno para la población blanca y otro para la negra. Las preguntas acerca del entorno eran casi idénticas para los dos casos. Las preguntas actitudinales eran también idénticas en ambos casos, pero un gran número de preguntas fueron dirigidas exclusivamente a un grupo racial o al otro. Los cuestionarios contenían preguntas actitudinales preguntando a los encuestados acerca de la satisfacción con los servicios del vecindario, sus sentimientos con respecto a la efectividad del gobierno para tratar con problemas urbanos, sus relaciones interraciales, sus actitudes hacia la integración y su percepción con respecto a la hostilidad entre las razas. El cuestionario en la Figura 11.4 es una versión acortada del cuestionario presentado a la población negra.

Usted notará que el cuestionario comienza con números de identificación para la persona siendo encuestada así como su localización. Hay también espacio para proveer información acerca de cuando comenzó la encuesta. La pregunta 1 es un ejemplo de una pregunta de actitud sobre el grado de satisfacción con los servicios brindados por la ciudad. La pregunta fue presentada en forma de matriz. Note también que las instrucciones están dirigidas a ambos, el encuestador y el encuestado.

La pregunta 2 tiene componentes cerrados y abiertos (A). El ítem A es también una pregunta de contingencia. La primer parte es la pregunta filtro, y la segunda es la pregunta de contingencia, que se aplica solo a los encuestados que hayan marcado categorías específicas en la primer parte. Todas las preguntas usan un código numérico, que es marcado por los encuestadores.

La sección final demuestra ventaja relativa de la encuesta personal sobre otros modos de llenar cuestionarios (por correo o teléfono). El encuestador puede proveer información detallada sobre la actitud general de los encuestados, que puede ayudar a interpretar su patrón de respuestas.

⁵⁷ Basado en *Racial Attitudes in Fifteen American Cities*, de Angus Campbell y Howard Schuman, (Ann Arbor, Mich.: Social Science Archive, 1973).

Figura 11.4

Cuestionario estudio problema urbano (versión resumida)

Tiempo de comienzo de la entrevista _____ a.m. p.m.	Número de ciudad <input type="checkbox"/> <input type="checkbox"/> v.3	SOLO PARA USO OFICIAL <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> V.2
	Número de segmento <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> v.9	
	Número de línea <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> v.10	
	Número de persona <input type="checkbox"/> v.19	

1. Primero, quisiera saber cuan satisfecho está usted con algunos de los principales servicios que la ciudad debe supuestamente proveer para su barrio. Acerca de la calidad de las escuelas públicas en su barrio, ¿está usted generalmente satisfecho, algo satisfecho o muy insatisfecho?

(CÓDIFIQUE A, Y PREGUNTE B HASTA E)

	Generalmente satisfecho	Algo satisfecho	Muy insatisfecho	No sabe
A. Calidad de las escuelas públicas	1	2	3	8
B. Parques y plazas para niños en este barrio	1	2	3	8
C. Deportes y centros de recreación para adolescentes en este barrio	1	2	3	8
D. Protección policial en este barrio	1	2	3	8
E. Recolección de basura en este barrio	1	2	3	8

2. Pensando en lo servicios de la ciudad, como escuelas, parques y recolección de basura, ¿usted cree que el barrio está mejorando, esta igual, o está empeorando sus servicios en comparación a otras partes de la ciudad?

Mejor..... (Pregunte A).....1
Igual.....2
Peor.....(Pregunte A).....3
No sabe.....8

A. SI RESPONDIÓ MEJOR O PEOR: ¿Cuál es la razón por la cual este bario mejora o empeora sus servicios)

3.Si usted tiene una queja seria los malos servicios de la ciudad, ¿piensa usted que puede lograr que los oficiales de la ciudad hagan algo al respecto si usted los llama?

Sí..... (Pregunte A).....1
No..... (Pregunte A).....5
No sabe.....(Pregunte A).....8

A. Ha llamado alguna vez a un oficial de la ciudad para quejarse por los pobres servicios?

Si.....1
No.....2

4. En general, usted piensa que los oficiales de la ciudad prestan más, menos o igual atención a un pedido o queja de una persona negra que a una queja de una persona blanca?

Mas..... 1
Menos2
Igual3
No sabe8

Ahora hablemos de los problemas de la ciudad en su conjunto.

5. ¿Cree usted que el alcalde de la ciudad esta intentando lo más fuertemente que puede de resolver los problemas de la ciudad, o que no está haciendo todo lo que puede para resolver los problemas?

Intentando lo más que puede1

No intentando lo más que puede (PREGUNTE A) X
 No sabe8

A. SI NO ESTÁ HACIENDO TODO LO QUE EL / ELLA PUEDE: Cree usted que están tratando medianamente fuerte para resolver estos problemas o no los está resolviendo para nada?
 Medianamente fuerte.....2
 Para nada.....3

LLENE LOS SIGUIENTES ÍTEMS INMEDIATAMENTE DESPUÉS DE DEJAR AL ENCUESTADO

<p>A. Tiempo total de la entrevista: minutos _____</p>	<p>E. ¿Cuáles personas mayores de 14 años estuvieron presentes durante la entrevista? CIRCULE TODAS LAS QUE CORRESPONDAN</p> <p>v.63 Ninguna.....0 Esposo/a.....1 Padre/s.....2 Niño mayor de 14.....3 Otro familiar o amigo.....4 Otros (ESPECIFIQUE).....5</p>
<p>B. Cooperatividad del encuestado Muy cooperativo.....1 Algo cooperativo.....2 Nada cooperativo.....3</p>	<p>F. Prolijidad del interior del hogar: V.64 Muy prolijo y limpio.....1 Medianamente prolijo y limpio....2 Medianamente desprolijo.....3 Muy desprolijo.....4</p>
<p>C. Interés del encuestado en cuestiones raciales: Gran interés.....1 Interés normal.....2 Poco interés.....3</p>	<p>G. Fecha de la entrevista: _____ V.69</p>
<p>D. Entendimiento del encuestado acerca de las diferentes cuestiones Buen entendimiento.....1 Mediano entendimiento.....2 Pobre entendimiento.....3</p>	<p>H. Firma del entrevistador: □□□ _____</p>

1. Por favor, de una breve descripción del entrevistado y de cualquier condición especial que pueda haber afectado la encuesta.

Resumen

1. El fundamento de todo cuestionario es la pregunta. El cuestionario debe traducir los objetivos de investigación en preguntas específicas. Las respuestas a estas preguntas proveerán los datos necesarios para probar las hipótesis.

2. La mayoría de las preguntas pueden ser clasificadas como preguntas fácticas o como preguntas sobre experiencias subjetivas. Las preguntas fácticas pueden ser diseñadas para obtener información acerca del encuestado. Las preguntas subjetivas tratan sobre las inclinaciones, preferencias, prejuicios, ideas, miedos y convicciones. En general, las preguntas subjetivas son de construcción mucho más compleja que las preguntas sobre datos personales. Las respuestas a estas preguntas son más sensibles a cambios en las palabras, énfasis y secuencia que las respuestas a preguntas fácticas.

3. Se puede distinguir entre cuatro tipos de estructura: preguntas abiertas, preguntas cerradas, preguntas de contingencia y preguntas de matriz.

En las preguntas cerradas se les ofrece a los encuestados un grupo de categorías de respuesta de las cuales deben escoger aquella que esté más cerca de representar sus opiniones. Las preguntas abiertas no son seguidas de ningún tipo de opciones y las respuestas de los encuestados son grabadas en forma completa. Una pregunta de contingencia se aplica solo a un sub grupo de encuestados. La relevancia de esta pregunta para es determinada por la respuesta de todos los encuestados a una pregunta filtro precedente. La pregunta matriz es un método para organizar un gran grupo de ítems que tienen las mismas categorías de respuesta.

4. Uno de los formatos más comunes para preguntas de cuestionarios es la escala de clasificación, a través de la cual los encuestados realizan juicios en cuanto a grupos de categorías ordenadas. Hay varios tipos de escalas de clasificación, incluyendo el ordenamiento de cartas y el diferencias semántico. El ordenamiento es utilizado en cuestionarios cuando el objetivo es obtener información acerca del grado de importancia de las prioridades que la gente aplica a un grupo de actitudes u objetos.

5. Las preguntas deben ser formuladas de modo tal que puedan ser comprendidas por todos los encuestados. Una pregunta tendenciosa está formulada de modo tal que parece al encuestado que el investigador espera una respuesta determinada. Las preguntas amenazadoras elevan el nivel de ansiedad de los encuestados. Ambos tipos de pregunta pueden conducir al sesgo en la respuesta. Las preguntas tendenciosas deberían ser evitadas, y las preguntas amenazadoras deben ser construidas con gran sensibilidad, utilizando técnicas especiales, tales como la utilización de una larga introducción a la pregunta, y la utilización de preguntas abiertas en lugar de preguntas cerradas.

Términos claves para revisión

Pregunta	Cuantificadores
Pregunta fáctica	Pregunta matriz
Actitud	Ordenamiento de cartas
Opinión	Diferencial semántico
Pregunta cerrada	Ordenamiento
Pregunta abierta	Pregunta tendenciosa
Pregunta de contingencia	Pregunta amenazadora
Pregunta filtro	Pregunta de opción múltiple
Clasificación	

Preguntas de estudio

1. Discuta las varias formas en las cuales las preguntas pueden utilizarse para obtener información fáctica, opiniones y actitudes de los encuestados.
2. Explique los usos de las preguntas abiertas, cerradas y de contingencia.
3. Enumere y describa los formatos utilizados para realizar preguntas con propósitos varios.
4. Discuta la importancia de la secuencia de las preguntas en un cuestionario.
5. Enumere los varios problemas que pueden surgir en la construcción de cuestionarios.

Capítulo 18

Construcción de Índices y Escalas

Las escalas (al igual que los índices) son instrumentos de medición⁵⁸. Ambos son contruidos para representar con mayor validez las complejidades inherentes al comportamiento humano. Conceptos tales como poder, equidad, libertad, inteligencia y burocracia son extremadamente difíciles de medir debido a que, entre otras cosas, son compuestos de diversas propiedades empíricas. Las escalas (y los índices) son medios para medir estos fenómenos complejos.

Las escalas y los índices son en la mayoría de los casos medidas compuestas contruidas sobre la base de la combinación de dos o más ítems o indicadores.

Ambos son empleados en las ciencias sociales por diversas razones:

1. Permiten la representación de varias variables por un único score que reduce la complejidad de los datos.
2. Proveen medidas cuantitativas que son susceptibles de volverse más precisas y de ser manipuladas con una mayor precisión estadística.
3. Aumentan la confianza de la medición. Un score en una escala o en un índice es un indicador más confiable de la propiedad bajo medición que una medida basada en una respuesta a una única pregunta o ítem en los cuales la probabilidad de cubrir el universo total de contenido es muy baja y un error en la interpretación en cualquiera de las preguntas podría llevar a una conclusión errónea acerca del fenómeno. Las escalas y los ítems se utilizan, entonces, para aumentar la confiabilidad de la medición y para obtener una mayor precisión.

Las escalas se diferencian de los índices por el mayor rigor en su construcción. Mientras que los índices son contruidos por la simple acumulación de scores, una mayor atención debe prestarse en el caso de las escalas para testear la validez y confiabilidad. Asimismo, la mayoría de las escalas suponen el principio subyacente de la unidimensionalidad. Este principio implica que los ítems comprendidos en la escala deberían reflejar una única dimensión y pertenecer a un continuum del cual se presume que refleja sólo un concepto.

Los análisis de ítems de algunas técnicas escalares identifican aquellas preguntas o ítems que no son parte del conjunto de ítems. Otras técnicas escalares permiten rankear ítems por su nivel de dificultad o intensidad. Inclusive, algunos métodos escalares producen escalas de nivel de medición intervalar, evitando consecuentemente las limitaciones impuestas por los datos nominales y ordinales.

La escala Likert mide actitudes en un nivel de medición ordinal e intervalar. La escala Guttman puede aplicarse a niveles nominales u ordinales de análisis.

Escala Likert

Es un método diseñado para medir actitudes personales. En su construcción pueden distinguirse seis pasos:

⁵⁸ Medición: procedimiento en el cual el investigador asigna símbolos o números a propiedades empíricas de acuerdo a determinadas reglas.

1. **Compilación de posibles ítems escalares.** Los investigadores deben compilar una serie de ítems que expresan un amplio rango de actitudes, desde las muy positivas a las muy negativas. Cada ítem ofrece cinco opciones pre-fijadas (MD, D, NA/ND, A, MA; Casi siempre, frecuentemente, ocasionalmente, raramente, casi nunca). En este continuum de cinco puntos los pesos 1, 2, 3, 4 y 5 o 5, 4, 3, 2 y 1 son asignados teniendo en cuenta el carácter positivo o negativo del ítem.
2. **Administración de los ítems a una muestra probabilística de encuestados.**
3. **Cálculo de un score total para cada encuestado.**
4. **Determinación del poder discriminante de los ítems.** El investigador tiene que determinar una base para la selección de los ítems que integrarán la escala final. Esto puede realizarse tanto con el método de la consistencia interna (correlacionando cada ítem con el score total y conservando aquellos con la correlación más alta) o con el método de análisis del ítem. Con ambos métodos el problema se encuentra en encontrar ítems que consistentemente separan los altos de los bajos. Esto se denomina poder discriminante (PD) de los ítems. En el cálculo del PD se suman los scores obtenidos en cada ítem por cada encuestado y se despliegan los scores en una matriz (*array*), usualmente del más bajo al más alto. El paso siguiente consiste en comparar el rango por sobre el primer cuartil (Q1, top 25%) con aquel por debajo del último cuartil (Q4) y el PD se calcula a partir de la diferencia entre las medias ponderadas (*weighted means*) de los scores por sobre el Q1 y aquellas que caen por debajo del Q4.

Tabla para el cálculo del PD de un (1) ítem

	N	1	2	3	4	5	Total ponderado	Media ponderada	PD Q1-Q4
Q1 Primer 25%	9	0	1	2	3	3	35	3.89	2
Q4 último 25%	9	1	8	0	0	0	17	1.89	

Total ponderado: score x frecuencia

Media ponderada: total ponderado / N

5. **Selección de los ítems.** El valor del PD se computa para cada uno de los posibles ítems y se seleccionan aquellos ítems con un mayores PD. Estos son los ítems que mejor discriminan los individuos que expresan diferentes actitudes hacia la actitud bajo medición.

6. **Testeo de la confiabilidad.** La confiabilidad de una escala puede ser testeada en una manera muy similar a la utilizada con otros procedimientos de medición. Por ejemplo, se puede seleccionar suficientes ítems para dos escalas (al menos 100) y dividirlos en dos sets, constituyendo dos escalas. El método de las dos mitades puede utilizarse a continuación (split-half reliability test).

La mayoría de los programas de computación incluyen procedimientos y estadísticas que facilitan la selección de ítems y la evaluación de cuán bien varios ítems miden el fenómeno subyacente.

Una de las medidas estadísticas más simple para examinar es la correlación bivariada de cada ítem con el total de la escala (R de Pearson). En general, los ítems que tienen una asociación fuerte con otros ítems mostrarán una correlación más alta con la escala. La examinación de la correlación bivariada ayuda a decidir cuáles ítems deben ser incluidos en la escala y cuáles deben ser descartado. Otra medida estadística de utilidad es el Alpha de Cronbach, que es una estimación del promedio de los coeficientes de confianza de todas las posibles mitades (*is an estimate of the average of all possible split-half reliability coefficients*, para una discusión acerca de la confianza ver Chapter 7). El alpha mide la extensión en la cual los ítems individuales comprendidos en la escala están correlacionados (*“hang together”*). Un alto alpha (0.70 es un nivel aceptable) indica que los ítems en la escala están “íntimamente conectados” (*“tightly connected”*).

Escala Guttman

Guttman sugiere que si los ítems comprendidos en la escala aluden (tap) a la misma dimensión actitudinal, estos pueden ser ordenados/arreglados (*arranged*) de manera tal de que haya un continuum que indique distintos grados de la dimensión subyacente. Las escalas Guttman son unidimensionales y acumulativas (*cumulative*). La “acumulatividad” implica que los ítems pueden ser ordenados por grado de dificultad y que los encuestados que responden positivamente a un ítem difícil también responderán positivamente a los ítems menos difíciles, o viceversa. Ej.: sabemos que si un padre hispanico no se molesta si su hija se casa con un afro-americano, tampoco se molestará si éste pertenece a su mismo club social o vive en su vecindario.

Tabla 18.6.
Una hipotética escala Guttman perfecta

Encuestado	Item 1: admite que se case con su hija	Item 2: admite que pertenezca al mismo club social	Item 3: admite que sea su vecino	Score total
A	+	+	+	3
B	-	+	+	2
C	-	-	+	1
D	-	-	-	0

Esta escala es unidimensional y acumulativa: los ítems están ranqueados unidimensionalmente en base a una única dimensión subyacente; ninguno de los encuestados tiene una respuesta negativa antes de una positiva o viceversa. Consecuentemente, la información referida a la posición de la última respuesta positiva de cualquier encuestado permite la predicción de todas sus respuestas a los otros ítems de la escala.

En la práctica, una escala Guttman perfecta se obtiene en raras ocasiones. En la mayoría de los casos se presentan inconsistencias. Consecuentemente, es necesario establecer un criterio para evaluar la asunción de unidimensionalidad y acumulatividad. Para ello Guttman desarrolló el Coeficiente de Reproducibilidad (Coefficient of Reproducibility, CR), que mide el grado de conformidad con un patrón escalable perfecto (*scalable pattern*):

$$CR : 1 - \frac{\sum e}{Nr}$$

Siendo:

$\sum e$: el número total de inconsistencias

N r: el número total de respuestas (número de casos x número de ítems)

Un CR de 0.90 es el standard mínimo para aceptar una escala como unidimensional.

Selección de ítems.

Gorten (1977:46) lista tres condiciones que deben cumplirse para descubrir y seleccionar ítems para una escala Guttman.

1. Debe haber una actitud hacia el objeto (clase de objetos, evento o idea) en la mente de las personas que conforman la población a ser muestreada y testeada.
2. El set de afirmaciones acerca del objeto debe tener significado para los miembros de la muestra y provocar en ellos una respuesta que sea un indicador válido de la actitud.
3. Los ítems en el set de afirmaciones o preguntas deben representar diferentes grados de una misma dimensión.

Los ítems de las escalas de actitud son seleccionados mediante una variedad de métodos de diferentes fuentes: diarios, libros, artículos académicos y el propio conocimiento del investigador acerca del problema. La realización de entrevistas a expertos y a un subgrupo de encuestados también ayudará a obtener buenos ítems. Una vez que un gran set de ítems es compilado debe realizarse una selección preliminar de los ítems. Los ítems seleccionados deben relacionarse claramente con la actitud bajo medición y cubrir el continuum total, desde las afirmaciones fuertemente favorables a las fuertemente desfavorables. Para cada afirmación deben construirse de dos a siete categorías de respuesta. Los formatos más comunes son los ítems al estilo Likert con escalas de cinco puntos.

Antes de otorgar puntajes (score) a las respuestas al cuestionario, los ítems deben ser ordenados de manera tal de que los números más altos serán otorgados a los ítems que reflejen los sentimientos más positivos o más negativos. Aquellos ítems que no correspondan a este patrón deberán ser invertidos/revocados (*reversed*).

Cálculo del Coeficiente de Reproducibilidad. Este coeficiente es definido como la extensión a la cual el patrón total de respuestas a un set de ítems puede ser reproducido si únicamente conocemos el score total. Esto depende de la extensión en la cual el patrón de respuesta se adecua a la patrón perfectamente escalable tal como se demuestra en la tabla 18.6. Cuando el coeficiente obtenido es menor al criterio requerido de 0.90, la escala debe ser refinada hasta que el CR alcance el nivel deseado.

Aplicación de la Escala Guttman: un ejemplo

Luego de que la escala Guttman ha sido desarrollada y refinada los resultados pueden presentarse para describir la distribución de la variable bajo medición. La escala puede también ser relacionada a otras variables en el estudio. Un ejemplo del desarrollo y aplicación de una escala Guttman es el estudio de Wanderer acerca de la severidad de los disturbios en ciudades americanas. Esta es una aplicación particularmente interesante de la técnica Guttman porque se basa en indicadores de comportamiento antes que en actitudes. La escala incluye los siguientes ítems de severidad de los disturbios: asesinato, llamado a la Guardia Nacional, llamado a la

policía estatal, disparos (sniping), saqueos (looting), interferencia con los bomberos (firefighters) y vandalismo. Estos ítems son ordenados de más a menos severos o de menos a más reportados. El coeficiente de reproductibilidad de esta escala Guttman de severidad de disturbios es de 0.92. En la tabla 18.7. se presenta la escala y la distribución de ciudades a lo largo de la escala. Las ciudades son organizadas en siete tipos de escala (*scale types*) de acuerdo con el grado de severidad, siendo 8 el menos severo y 1 el más severo.

Tabla 18.7
Una escala Guttman de Severidad de Disturbios

Tipo de la escala	% de ciudades	Ítems reportados
8	4	Ningún ítem de la escala
7	19	"Vandalismo"
6	13	Todos los anteriores e "interferencia con los bomberos"
5	16	Todos los anteriores y "saqueo"
4	13	Todos los anteriores y "disparos"
3	7	Todos los anteriores y "llamado a la policía estatal"
2	17	Todos los anteriores y "llamado a la Guardia Nacional"
1	11	Todos los anteriores y "oficial de la ley o civil asesinado"
total	100	

En un segundo momento, el investigador trató a la severidad de los disturbios medida por la escala Guttman como una variable dependiente y examinó un set de variables independientes en términos de su relación con la severidad de los disturbios. Por ejemplo, se encontró una relación entre el porcentaje de no blancos involucrados en el disturbio y la severidad del disturbio. Es decir, que una vez que un disturbio tiene lugar, a mayor porcentaje de no blancos involucrados, mayor severidad del disturbio.