

CONSTRUCCIÓN Y ANÁLISIS DE DATOS

racterística subyacente²¹. Un **indicador** es una característica o respuesta manifiesta que se encuentra ya definida en términos de lo que hay que observar. La medida compleja que se obtiene combinando los valores obtenidos por un individuo en cada uno de los indicadores propuestos para la medición de una variable se denomina **índice**. "La diferencia entre un índice y un indicador es entonces de grado. Un índice es un complejo de indicadores de dimensiones de una variable, y constituye por lo tanto el indicador total de una variable compleja" (Korn, 1969: 12).

2. ESCALAS DE ACTITUDES

Si un índice no es más que una forma de reducir un espacio de propiedades, podemos decir del mismo modo que las escalas son una clase particular de índices²², cuya diferencia específica consiste en la utilización de una serie de procedimientos objetivos para la selección de los indicadores estandarizados si éstos se inscriben efectivamente dentro de una dimensión común.

²¹ Aunque juega aquí una suerte de "ley de rendimientos decrecientes". Según H. W. SMITH, "Los primeros ítems que se agregan contribuyen en mucho mayor medida a la confiabilidad que los últimos añadidos. Así, luego de un cierto punto (i.e. 8 a 10 ítems) cualquier ítem adicional contribuirá casi insignificativamente a aumentar la confiabilidad... ¿Por qué se mantiene esta relación entre la longitud de la escala y la confiabilidad? Cualquier medida tiene fuentes idiosincráticas de error. Por ejemplo, el uso del ingreso como una medida del *status* socioeconómico presenta problemas idiosincráticos como el de incluir en la clase alta miembros de la Mafia con un salario de 100.000 US\$. Con cada ítem agregado a la escala puede esperarse que decrezcan los errores ítem-específicos, puesto que normalmente se espera que se cancelen unos a otros. Esto es, puede esperarse que cada medida incluya factores irrelevantes para lo que el investigador está midiendo, pero como estos componentes irrelevantes son presumiblemente distintos para cada ítem, y como todos los ítems comparten un foco común, el conjunto total de medidas proveerá normalmente una medición más pura del fenómeno en estudio que cualquier medición aislada" (cf. SMITH, 1975: 60-61).

²² Según SAMUEL STOUTER, "Se dice que existe escala cuando a partir de una distribución de frecuencias multivariada de un universo de atributos, es posible derivar una variable cuantitativa con la cual caracterizar los objetos de un modo tal que cada atributo sea una función simple de aquella variable cuantitativa" (citado en PADUA, 1979: 160).

Las escalas fueron desarrolladas originalmente con miras a la medición de actitudes, aun cuando puedan ser aplicables a otra clase de objetos. En la *Enciclopedia Internacional de las Ciencias Sociales* (I, 15) podemos leer esta definición de Milton Rokeach: se entiende por "actitud" "una organización relativamente duradera de creencias en torno a un objeto o situación, que predispone a reaccionar preferentemente de una manera determinada".

Hay varios tipos de escalas de actitudes, aunque todos ellos se orientan a conocer la **dirección** (positiva o negativa) y la **intensidad** (mayor o menor)²³ de las actitudes de personas individuales hacia un determinado objeto utilizando un conjunto de ítems. Cada ítem consiste en un juicio, en una frase que funciona como indicador de la variable que se pretende medir. Así, el enunciado "Las mujeres no deberían meterse en política" puede funcionar como un ítem dirigido a medir la actitud hacia la participación de las mujeres en la política; o bien "las representaciones estudiantiles deberían participar en las decisiones sobre los planes de estudio", sería un ítem que mediría la actitud hacia la participación de los estudiantes en el gobierno universitario. De acuerdo con el modo en que es construida la escala y al tipo de análisis al que da lugar, es corriente distinguir entre tres tipos de escalas:

- centradas en el **sujeto**: se persigue ubicar al sujeto en el continuo actitudinal (típicamente, es el caso de las escalas Lickert);
- centradas en el **estímulo**: se basan en adjudicar un valor a cada estímulo (o ítem), por lo que la estrategia consiste en localizar los ítems en el continuo (son las de Thurstone, entre las que se cuenta su "escala de intervalos aparentemente iguales");
- centradas en las **respuestas**: aquí el propósito es escalar tanto a los sujetos como a los ítems, lo que da lugar a una escala acumulativa (es el caso del análisis de escalograma de Guttman).

Lo que diferencia estos tres tipos de escalas no son los ítems considerados en sí mismos -éstos pueden ser los mismos-, sino la relación lógica que guardan entre ellos. Por lo demás, las tres formas son semejantes en cuanto suponen en

²³ Puesto que una actitud neutral (ni positiva, ni negativa) se asocia generalmente a una baja intensidad, estas dos propiedades pueden combinarse en una única escala que mide simultáneamente dirección e intensidad.

Edwards (1957) ha formulado una serie de criterios²³:

- a. evitar ítems que apunten al pasado en vez del presente;
- b. evitar ítems que dan demasiada información sobre hechos, o aquellos que pueden ser interpretados como tales;
- c. evitar los ítems ambiguos;
- d. evitar los ítems irrelevantes con respecto a la actitud a medir;
- e. los ítems deben expresar actitudes favorables o desfavorables con respecto a la actitud, y no expresar distintos puntos del continuo;
- f. evitar los ítems con los cuales todos o casi nadie concuerda;
- g. los ítems deben ser formulados en lenguaje simple, claro y directo;
- h. preferentemente, no exceder las 20 palabras por ítem;
- i. un ítem debe contener sólo una frase lógica;
- j. evitar el uso de palabras como "todos", "siempre", "nadie", etc;
- k. de ser posible, los ítems deben ser formulados con frases simples, y no compuestas;
- l. usar palabras que el entrevistado pueda comprender;
- m. evitar las negaciones, en particular las dobles negaciones;
- n. combinar los ítems formulados positiva y negativamente en una proporción de 50%-50%.

Aunque a veces obvias y algo redundantes, estas recomendaciones constituyen una buena base para evaluar críticamente cada ítem. Sin embargo, la experiencia muestra que no toman en cuenta un aspecto esencial como es el grado de competencia lingüística de la población a encuestar. Es un error presumir que cualquier persona está en condiciones de responder coherentemente a una batería de ítems. En sectores con muy bajo nivel de educación formal, este tipo de ítems simplemente no funciona, o bien demanda la reformulación sistemática de los enunciados por parte del encuestador, con lo que las respuestas pierden comparabilidad ya que lo son a estímulos diversos.

Cada ítem consiste en un juicio (o enunciado) sobre el cual el juez -o el sujeto- deberá expresar su grado de acuerdo

²³ Los ítems reproducidos a partir del texto de PAVVA (1979).

principio el uso de jueces, miden en el nivel ordinal y se plantean como unidimensionales²⁴.

La escala Likert

Nos extendaremos exclusivamente sobre ésta, por ser el tipo más simple en cuanto a su construcción y en cuanto a su flexibilidad, que la hace adaptable a propósitos variados. La escala Likert consiste en un mero índice sumatorio. Se solicita la reacción del sujeto a una serie de ítems, cada uno de los cuales expresa tentativamente la actitud subyacente, y a los que el sujeto responde en términos de su grado de acuerdo-desacuerdo. Para esta escala, los ítems no deben escalararse a lo largo del continuo, sino que el método de construcción supone seleccionar ítems que sean definitivamente favorables o desfavorables con relación al objeto que se quiere medir. Sumando los puntajes obtenidos por un sujeto en cada ítem, le correspondrá un puntaje global que estará expresando su posición en el continuo. Los pasos a seguir en la construcción de la escala son los detallados a continuación.

1. La construcción de los ítems: se deben construir entre 30 y 50 ítems. Se trata esencialmente de un proceso imaginativo, construyendo ítems que resulten satisfactorios en términos de su validez facial o aparente por su relación con la característica que se pretende medir²⁵. Para la construcción de cada ítem,

²⁴ Sobre las escalas tipo THURSTONE Y GERTMAN, el lector podrá remitirse al trabajo bastante completo de PAVVA Y ANAYAS (cf. PAVVA, 1979). En lo que respecta específicamente al análisis de escalogramas de GUTMAN, se encontrará un excelente ejemplo completamente desarrollado en MARINKAS (1976). Para un repertorio bastante completo de índices y escalas, cf. MILLER (1991).

²⁵ Los ítems son construidos a partir de juicios que expresan una relación postulada a nivel de la teoría sustantiva, y de observaciones empíricas de grupos o sujetos que pertenecen a grupos o asociaciones que manifiestan la propiedad que se quiere medir. Los ítems así pueden ser extraídos de libros, publicaciones y artículos que tratan realmente sobre el objeto que se quiere medir. También puede el investigador acudir al análisis de contenido sobre discursos o manifiestos de individuos y asociaciones (por ejemplo, si se trata de medir radicalismo/conservadurismo, una fuente muy rica para la formulación de ítems son las declaraciones de grupos de interés: empresarios, grupos políticos de izquierda y de derecha, etc.). (cf. PAVVA, 1979: 170).

o desacuerdo, para lo que se le ofrece un número de alternativas que puede ser de 3, 4, 5, 6 ó 7.

Ejemplos de alternativas de respuesta

3 alternativas	5 alternativas	6 alternativas
<input type="checkbox"/> De acuerdo	<input type="checkbox"/> Muy de acuerdo	<input type="checkbox"/> Muy de acuerdo
<input type="checkbox"/> Ni de acuerdo ni en desacuerdo	<input type="checkbox"/> De acuerdo	<input type="checkbox"/> De acuerdo
<input type="checkbox"/> En desacuerdo	<input type="checkbox"/> Ni de acuerdo ni en desacuerdo	<input type="checkbox"/> Más bien de acuerdo
	<input type="checkbox"/> En desacuerdo	<input type="checkbox"/> Más bien en desacuerdo
	<input type="checkbox"/> Muy en desacuerdo	<input type="checkbox"/> En desacuerdo
		<input type="checkbox"/> Muy en desacuerdo

Algunos autores recomiendan incluir la alternativa "Sin opinión". Así, es sabido que muchas más personas contestarán "No sé" si esta alternativa les es explícitamente ofrecida que si no lo es. El problema de no separar los "sin opinión" es que pueden terminar confundándose con una actitud neutra, cuando se trata de casos distintos²⁷.

Es recomendable que los ítems sean precedidos de un breve texto del tipo:

"Las siguientes afirmaciones son opiniones con respecto a las cuales algunas personas están de acuerdo y otras en desacuerdo. Indique por favor (marcando con una x en el casillero correspondiente) la alternativa que más se asemeja a su opinión".

2. Administración de los ítems a una muestra: una vez contruidos los ítems, se los debe aplicar a la población-objeto. Si se desea validar previamente la escala puede aplicársela primero a una muestra de 50 a 100 personas que actuarán como jueces (idealmente estas personas deberán ser seleccionadas al azar de una población que presente ca-

²⁷ Se ha señalado, además, cómo "el tamaño de este efecto es parcialmente una función del modo en que la opción "sin opinión" es ofrecida", siendo el efecto de filtro mayor cuando la opción es dada mediante una pregunta previa y no sólo como una alternativa de respuesta (cf. COXETER y PRISNER, 1986: 35).

racterísticas similares a las que se desea estudiar).

3. Asignación de puntajes a los ítems: aquí comienza propiamente hablando el análisis de la escala. Cada ítem habrá sido clasificado *a priori* en positivo o negativo. Un ítem **positivo** es aquel que denotará una actitud favorable al objeto por parte de la persona que concuerde con él; por ejemplo, el ítem A6: "Los niños con SIDA deben ir a la misma escuela que los chicos sanos". En cambio, si alguien manifiesta su acuerdo con el enunciado del A3 -"Se debe prohibir el casamiento a las personas, que estén enfermas de SIDA"- nos inclinaremos a pensar que su actitud hacia los portadores de SIDA es más bien desfavorable, por lo que este será considerado un ítem **negativo**²⁸.

Si, por ejemplo, se ha optado por tres alternativas de respuesta, en un ítem positivo se asignarán 2 puntos a la persona que esté "de acuerdo", 1 a la respuesta neutra, y 0 a "en desacuerdo"; los ítems negativos serán en cambio puntuados de modo inverso (0 para "de acuerdo", 2 para "en desacuerdo"). La ventaja de contar con ítems positivos y negativos es controlar los efectos del *response-set*, es decir mantener controlados a los respondedores que tienden a contestar automáticamente por sí (*yeah-sayers*) o por no.

4. Asignación de puntajes totales para cada sujeto: se suman los puntajes obtenidos por cada encuestado. Asumiendo la hipótesis de que los ítems han sido bien elegidos -esto es, que están expresando efectivamente la actitud subyacente- se espera que las personas con actitudes favorables presenten puntajes totales elevados, mientras que las desfavorables obtendrán bajos puntajes. Así, por ejemplo, si la versión original de la escala consta de 30 ítems puntuados desde 2 hasta 0, los puntajes totales de cada sujeto podrán variar entre 0 y 60 puntos ($60 = 30 \times 2$).

²⁸ Estos ejemplos están tomados de una investigación realizada durante 1992 por estudiantes de Antropología Social de la UNaM.

grupos se calcula el puntaje promedio en cada ítem, pudiendo compararse luego la diferencia entre las medias. Esta es la base para analizar el poder discriminatorio de cada ítem: utilizan- do, por ejemplo, la t de Student, se pueden eliminar los ítems en los cuales la diferencia de las medias del grupo alto y del bajo no arrojan valores de t significativos. El procedimiento está basado en asumir la hipótesis de que en su conjunto los ítems están relacionados con la actitud que se persigue medir, de modo que se pueda eliminar a los ítems cuyos valores de respuesta no sean consistentes con los puntajes totales.

Así, en el ejemplo, se concluye fácilmente que el ítem A1 ("Conviene que antes de casarse las personas se hagan una prueba para saber si están enfermas de SIDA") no discrimina: todos los jueces han respondido del mismo modo, por lo que la media es idéntica en ambos grupos y la diferencia es 0. Otros ítems malos son el A12 ("Los trabajadores del hospital tienen derecho a negarse a atender a los enfermos de SIDA") y el A27 ("El SIDA es un producto del libertinaje sexual"), en los que las diferencias entre las medias son mínimas y, consecuen- temente, los valores de t no alcanzan un nivel de significación aceptable. De los ítems incluidos en la Tabla el mejor es el A10 ("Una empresa debe tener derecho a despedir a un em-pleado enfermo de SIDA"): la diferencia entre las medias es de 1,8 - sobre un máximo posible de 2-, y el valor de t -"9"- es significativo al nivel de 0,0001.

6. La versión final de la escala: en definitiva, se conser- van para la versión final únicamente los ítems cuya relación con el conjunto de la escala original se haya comprobado me- diante algún test de significación, y se descarta por completo el resto. La cantidad de ítems seleccionados según su poder discriminatorio (como mínimo, seis) constituirá la escala fi- nal. Esta se podrá aplicar entonces a la población-objeto, en el caso de haber trabajado con una muestra de jueces, o si no simplemente se computará el puntaje definitivo de los sujetos en los ítems que integran la versión final de la escala.

El puntaje de cada individuo en la escala será igual a la suma de los puntajes parciales obtenidos en cada ítem. Así, si

Tabla 5.1: Ejemplo de matriz para el análisis de ítems

Ítem	Jueces												Mediana	Mediana	Difer.	valor t	valor p					
	1	2	3	4	5	6	7	8	9	10	10	11						12	12	12	12	12
A1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	0	0	0	0,000	>0,9999
A3	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	0	0	0	0,0005	0,9995
A6	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	0	0	0	5,657	0,0001
A10	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	0	0	0	9,000	>0,0001
A12	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	0	0	0	1,000	0,3466
A24	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	0	0	0	2,236	0,0558
A27	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	0	0	0	-0,426	0,6811
Puntos	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	0,4	0,6	-0,2		
	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	2	1,8	1,8	1	

5. Análisis de los ítems: se ordenan todos los sujetos se- gún su puntaje, de mayor a menor; hecho lo cual se descarta el 50% central de la distribución para continuar trabajando sola- mente con los cuartiles 1 y 4. Así, sobre los veinte encuestados de la Tabla 5.3²⁷, nos quedaríamos sólo con diez: cinco consti- tuirán el grupo "Alto" (los 5 que presentaron los más altos

²⁷ Esta tabla fue confeccionada con fines puramente ilustrativos; son muy pocos sujetos e ítems. En realidad, se trabajó con una muestra de 140 personas de Posadas.

se trata de una escala de 10 ítems puntuados de 0 a 2, los puntajes de los individuos estarán comprendidos en el rango de 0 a 20 puntos. Estos puntajes extremos expresan claramente actitudes desfavorables y favorables, respectivamente. Pero la interpretación de los puntajes individuales que caigan entre los dos valores extremos es problemática, y no puede hacerse en términos absolutos. Dado que la escala mide en un nivel ordinal, y que la localización del punto neutro no es conocida, la interpretación de los puntajes intermedios sólo es posible en términos relativos al grupo estudiado³⁰.

3. CONFIABILIDAD Y VALIDEZ

El objetivo de las técnicas de recolección de datos (cuestionarios, tests, escalas, etc.) es obtener evidencia comprobada que sea significativa para responder a las preguntas que formulamos en nuestra investigación. Hemos visto que para ello debemos producir definiciones operacionales que nos permitan medir los conceptos en los que estamos interesados.

Cuando se trata de evaluar la bondad de cualquier instrumento de medición, el problema básico es definir cuáles deberán ser consideradas como verdaderas diferencias en la característica que se está midiendo distinguiéndolas de las variaciones que devienen de meros errores en la medición. Así, cuando me pregunto: "¿Qué es lo que mide mi instrumento?"; "¿Son relevantes los datos que me proporciona para la característica que busco medir?"; o bien "¿Las diferencias que resultan de la aplicación de mi instrumento representan verdaderas diferencias en la característica a medir, o están reflejando la incidencia de otros factores?", estoy inquiriendo por la validez³¹.

Por otra parte, la confiabilidad se refiere a la consistencia entre medidas independientes del mismo fenómeno. En este

caso, la pregunta es si el mismo instrumento utilizado en diferentes ocasiones y/o por diferentes investigadores producirá los mismos resultados³². Validez y confiabilidad aparecen por lo tanto como dos requisitos que deberá satisfacer cualquier instrumento de medición satisfactorio. Abordaremos sucesivamente estos dos aspectos, para luego concluir acerca de la relación que mantienen entre sí.

Validez

Desde Louis Guttman en adelante, ha sido usual dividir la cuestión en validez "interna" y "externa". Según Zetterberg: "La diferencia principal es que la primera expresa una relación 'lógica', mientras que la segunda expresa una relación 'empírica'. En otras palabras, la validez interna se puede apreciar sin estudios empíricos, mientras que la determinación de la validez externa consiste en poner a prueba una hipótesis" (1968: 92).

Validez interna: para referirse a este tipo, suelen emplearse de manera indistinta las expresiones "validez lógica" o "facial"³³, así como la de "validez de contenido"³⁴.

Aunque es evidente que la validez interna no es suficiente para evaluar la calidad de un instrumento, no deja de ser una condición necesaria, que debe ser satisfecha previamente a cualquier otro análisis de validez³⁵. Al decir de Goode y Hatt: "Se refiere al análisis, teórico o de 'sentido común'... y se emplea casi siempre debido a que dimana de la definición precisa del continuo" (1969: 292).

Retornemos al caso que presentábamos acerca de una medida de la idea de "satisfacción con el trabajo". Justamente Zetterberg ha ilustrado este tema recurriendo a ese mismo ejemplo (1968: 92 y s.). Hablar de "validez perfecta" impli-

³⁰ "No existe evidencia que sugiera que el punto neutro se corresponda con el punto medio de los valores escalares" dicen McIVER y CARMINE (1986: 28). Así, de una persona con 10 puntos en nuestro ejemplo no se puede afirmar que necesariamente tenga una actitud neutra hacia los enfermos de SIDA.

³¹ Es en este sentido que GOODE y HATT definen sintéticamente: "Una escala tiene validez cuando verdaderamente mide lo que afirma que puede medir" (1969: 292).

³² "Cabe confiar en una escala cuando produzca constantemente los mismos resultados al aplicarla a una misma muestra" (cf. GOODE y HATT, 1969: 289).

³³ Cf. GOODE y HATT, 1969: 292.

³⁴ Cf. CARMINE y ZELLER, 1979: 20.

³⁵ De lo contrario, se incurriría en el riesgo cierto de no saber -literalmente- de qué se está hablando.

