

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	Ítem 1: admite que se case con su hija	Ítem 2: admite que pertenezca al mismo club social	Ítem 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

Nr: 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.

