

Métodos Cuantitativos

Módulo 4: La medición

4.01 La medición: operacionalización

Antes de ocuparnos del tema de la medición, clarificaré brevemente los términos *variable* y *operacionalización*. Antes nos hemos referido a las variables como constructos operacionalizados, pero el término **variable** también se puede referir a la representación de un constructo que todavía es en cierto modo abstracto. Utilizamos el término **operacionalización** si queremos indicar explícitamente que estamos hablando sobre un método específico, concreto, para medir o manipular un constructo.

Digamos que estoy interesada en el constructo “poder político”. Puedo representar este constructo con la variable “influencia en el parlamento”. Esta representación es más específica, pero todavía se nos pueden ocurrir varios procedimientos para operacionalizar o medir la variable “influencia en el parlamento”.

Por ejemplo, yo puedo contar el número de propuestas presentadas por un miembro del parlamento o por el Congreso, o puedo contar el número de años que alguien ha estado en el parlamento, o puedo pedirles a analistas políticos que califiquen los miembros del parlamento en términos de su influencia.

De forma que operacionalización significa selección o creación de procedimientos específicos para medir o manipular el constructo de nuestro interés. Una operacionalización hace posible asignar a las personas puntajes reales respecto a la variable de interés.

Supongamos que quiero operacionalizar el constructo “amor por los animales”. Puedo observar a una persona interactuando con un gato y contar qué tan seguido la persona acaricia o golpea al gato. También puedo decidirme por una operacionalización diferente o **definición operacional**, creando un cuestionario con enunciados como “Usualmente evado a las mascotas de las demás personas” y “Yo amo a los animales”.

¿Qué pasa con las variables independientes que son manipuladas? Bueno, supongamos que yo quiero saber si la exposición a los animales incrementa el amor por los animales. Puedo operacionalizar la variable “exposición a los animales” creando dos niveles de exposición. Puedo asignar aleatoriamente a personas para cuiden de un gato de un refugio durante un mes, o asignarlos a la condición de control, esto es, sin gato.

Otra operacionalización puede ser llevar la mitad de un aula escolar a un zoológico interactivo y la otra mitad a un museo abstracto. O puedo asignar a los participantes para que observen un documental sobre animales y un documental sobre trenes. Como pueden ver, las posibilidades de operacionalización son innumerables. Por supuesto algunas operacionalizaciones son mejores que otras.

Una operacionalización no necesariamente captura o representa el constructo en su totalidad. A medida que nuestros constructos son más “difusos” o más complejos, hay una mayor probabilidad de que midamos o manipulemos sólo una parte del constructo. Por ejemplo, si medimos el amor por los animales con un cuestionario de autoevaluación, medimos sentimientos y actitudes, lo cual ofrece una imagen más positiva. Si medimos el amor por los animales colocando cámaras en los hogares de las personas y observando su comportamiento, podemos encontrar puntajes más bajos.

Podemos encontrar que, comparado con la autoevaluación, la gente muestra mucho menos amor cuando su gato los despierta a las 5 AM o bloquea la vista del televisor cuando están viendo su programa favorito. Es importante tener presente qué aspecto del constructo es el que realmente la operacionalización va a medir o a manipular, especialmente cuando usamos datos para llegar a conclusiones acerca del constructo entero de nuestra hipótesis. Nuestra conclusión puede aplicarse sólo a un aspecto limitado del constructo.

4.02 La medición: estructura de la medición

Una vez que hemos operacionalizado el constructo estamos listos para empezar a medir. Infelizmente, la medición es mucho menos clara y directa en las ciencias sociales que en las ciencias naturales. Por lo tanto, es extremadamente importante saber a qué nos referimos cuando decimos que estamos *midiendo* la depresión o la persuasión política. Deberíamos saber que qué estamos capturando con los números que resultan de la medición, pero más importante que eso, deberíamos saber qué información *no* estamos capturando.

¿Entonces qué es la medición? Ok. Aquí vamos: la medición es la representación de las relaciones entre objetos, personas o grupos en una cierta propiedad¹, utilizando relaciones entre números. Tomemos, por ejemplo, la longitud corporal. Puedo determinar relaciones cualitativas en esta propiedad entre tres “objetos” (en este ejemplo: dos personas y un gato) colocándolos uno al lado del otro, observando y comparando qué tanto sobresalen sus cabezas.

La primera cosa que puedo decir es que ellos son todos diferentes en cuanto a longitud. Puedo representar esta relación desigual utilizando diferentes etiquetas o números para representar las desigualdades. Por supuesto sería raro utilizar los números 2, 10 y 14, porque hay otro tipo de relación que podemos ver inmediatamente, lo cual *no* se refleja en los números asignados. Me refiero a la relación de orden entre A, B, y C. La persona A es la más alta, así que debe recibir el número mayor, 14; la persona C es la más corta y debería recibir el número más bajo, 2. Utilizamos este orden de los números para representar el orden de las personas –y el gato– en términos de longitud corporal.

Y no tenemos por qué parar ahí. También podemos determinar si la diferencia en la longitud entre la persona A y la persona B es la misma o mayor que la diferencia entre la persona A y la persona C. Sostenemos un pedazo de cartón sobre la cabeza de A y cortamos el cartón cuando alcanza la parte de arriba de la cabeza de B. Luego sostenemos el pedazo de cartón sobre la cabeza de C para

¹ “Propiedad”, del inglés “property”, en el sentido de “atributo” o “característica”. Nota de la traductora.

comparar con A. Supongamos que el cartón alcanza exactamente hasta la cabeza de A, entonces las diferencias en longitud son las mismas.

Podemos representar esta relación de “diferencias iguales” utilizando números que difieren por la misma cantidad. Por ejemplo, la diferencia entre los números para B y A, 14 y 10, es 4, así que podemos cambiar el número asignado a C de 2 a 6. Las diferencias iguales de números entre B y A, $14-10=4$ y entre A y C, $10-6=4$, ahora reflejan con exactitud las diferencias iguales en la longitud corporal entre A y B y A y C. si la diferencia entre A y C ha sido mayor que la diferencia entre A y B, entonces la diferencia en los números correspondientes debe haber reflejado esto.

Hay un tipo más de relación que podemos observar para la propiedad longitud corporal. Podemos comparar la proporciones² de la longitud corporal. Podemos tomar un pedazo de cartón que cortamos previamente, cortar algunos pedazos extra de exactamente la misma longitud y cuántas unidades de cartón representan la altura de C, A y B.

Ok, ahora supongamos que se necesitan dos unidades de cartón para alcanzar la cima de la cabeza de C, cuatro para alanzar la cima de la cabeza de B y tres para alcanzar la cima de la cabeza de A. Esto significa que la altura de B es dos veces la de C. Podemos reflejar esta relación cambiando los números otra vez a 9, 12 y 6. Todavía estamos utilizando números diferentes para longitudes diferentes, el orden de los números corresponde al orden de las longitudes corporales, las diferencias entre A y B y entre A y C son las mismas ($12-9=3$ y $9-6=3$) y ahora el número para la persona B es dos veces mayor que el número para C.

Ahora podemos ver que la medición es la representación de relaciones empíricas, entre objetos o personas en cuanto a una cierta propiedad, utilizando las relaciones numéricas entre números. Podemos diferenciar longitudes, ordenarlas y comparar diferencias y proporciones de longitud corporal. Determinamos estas relaciones empíricas observando y utilizando unas piezas de cartón. Por supuesto este método es muy engorroso si queremos evaluar la longitud corporal de un gran número de personas.

Asignar números de manera que capturen estas relaciones empíricas, por ejemplo, utilizando una cinta métrica, hace nuestra vida mucho más fácil; especialmente si queremos comparar o agrupar mucha gente. Y por supuesto asignar números para representar una propiedad nos permite utilizar la estadística para ayudar a describir y a extraer conclusiones sobre la propiedad en la cual estamos interesados.

4.03 La medición: niveles de medición

Vimos que la medición es la representación de relaciones entre personas en cuanto a cierta propiedad utilizando relaciones correspondientes entre números. Consideremos la longitud corporal. Para esta propiedad podemos distinguir entre diferentes longitudes corporales, podemos ordenarlas, podemos comparar diferencias e incluso comparar proporciones. Podemos usar las relaciones numéricas entre cifras para representar todas estas relaciones.

² Del inglés “ratio”: razón, proporción, cociente. Nota de la traductora.

Para la longitud corporal podemos interpretar significativamente las cuatro relaciones posibles: desigualdad, orden, diferencias y proporciones. Infelizmente, este no es el caso para propiedades psicológicas y sociales. Para muchas propiedades en las ciencias sociales podemos sólo determinar algunas de esas relaciones. El término **nivel de medición** es utilizado para indicar qué tipo de relación puede ser significativamente interpretada.

Si la única relación que podemos determinar es la de la desigualdad, distinguiendo entre valores, entonces llamamos a esto una **variable nominal**. De un instrumento que sólo puede diferenciar entre valores se dice que tiene un **nivel nominal de medición**. Ejemplos de esto son la nacionalidad, el sexo o la preferencia de mascotas.

Un alemán tiene una nacionalidad diferente a la de un británico, las mujeres son de un sexo diferente al de los hombres, las personas a las que les gustan los perros tienen una preferencia de mascotas diferente a las de las personas que gustan de los gatos o de los hamsters. Un valor no representa un mayor grado de la propiedad estudiada que cualquier otro valor, simplemente son diferentes. Un alemán no tiene más nacionalidad que un británico, una mujer no tiene más sexo... bueno... digamos más género que un hombre y que a uno le gusten los gatos no significa que uno tenga una preferencia más fuerte por los animales que alguien a quien le gusten los gatos o los hamsters. No hay un orden entre esas propiedades.

Las variables **ordinales** permiten la diferenciación y la ordenación de los valores. Supongamos que yo quiero medir la habilidad matemática y utilizar para eso el número de respuestas correctas en un test de matemáticas de 10 preguntas. Cuanto mayor sea la habilidad matemática de alguien, más preguntas correctas tiene. Podemos ordenar los puntajes obtenidos por las personas para reflejar su orden en habilidad matemática, pero las diferencias o proporciones de los puntajes no reflejan las diferencias o proporciones en habilidad matemática.

No tenemos manera de mostrar que la diferencia entre un puntaje de cuatro y uno de cinco es la misma (en términos de diferente habilidad matemática) que la que existe entre un puntaje de siete y uno de ocho. Seguramente la diferencia en respuestas correctas es la misma, pero ¿cómo podemos mostrar que esto corresponde a una igual diferencia en habilidad matemática? No podemos. Y lo mismo pasa con las proporciones: alguien con un puntaje de 10 no tiene el doble de habilidad matemática que alguien con un puntaje de cinco.

En realidad, queda por verse si los puntajes de los tests son medidos a nivel ordinal, esto es, si reflejan con exactitud el orden en habilidad matemática. ¿Qué pasa si alguien con un puntaje de uno se pasó todo el tiempo en la pregunta más difícil y la resolvió correctamente, mientras que alguien más se concentró en las preguntas más fáciles y tuvo un puntaje de tres?

Sólo si las preguntas son igualmente difíciles, podemos utilizar los puntajes de los tests para reflejar con precisión el orden de los estudiantes en habilidad matemática. En ese caso se dice que el test de matemáticas mide a **nivel ordinal**.

Para **variables de intervalo** es posible distinguir y ordenar los valores, pero también interpretar las *diferencias* entre valores. La temperatura es un buen ejemplo. Supongamos que estoy calentando cuatro recipientes llenos de agua en la estufa y que mido la temperatura con un termómetro en

grados Fahrenheit. Un recipiente de agua que da 90 grados Fahrenheit está más caliente que uno que da 80; podemos verificar introduciendo nuestra mano. Y lo mismo aplica para dos recipientes que dan una medición de 40 y 50 grados.

También podemos verificar que cuando subimos la temperatura del agua de 80 grados hasta 90 grados, la expansión del líquido, como el mercurio en un termómetro, es la misma expansión que cuando subimos la temperatura del agua a 40 grados hasta 50 grados. Así que la diferencia entre 80 y 90 y entre 40 y 50 es la misma.

No podemos decir, sin embargo, que el agua a 80 grados Fahrenheit es dos veces más caliente que el agua a 40 grados. Esto, porque el punto cero para la temperatura es definido arbitrariamente. El valor cero no corresponde a la ausencia de temperatura, corresponde a la temperatura requerida para congelar el agua salada. La escala Celsius define el cero como la temperatura a la cual se congela el agua dulce. Si consideramos las mismas temperaturas de antes, pero ahora en grados Celsius veremos que treinta y dos punto dos menos veintiséis punto seis es cinco punto seis, del mismo modo que diez menos cuatro punto cuatro es cinco punto seis. Pero veintiséis punto seis no está ni cerca de cuatro punto cuatro. Esto es así porque las escalas utilizan diferentes puntos cero.

A diferencia de los intervalos variables, las variables de **proporción** tienen un punto cero no-arbitrario que es el mismo para cualquier escala que uno escoja. Por supuesto la longitud es un ejemplo obvio. La ausencia de longitud, longitud 0, es la misma sea que midamos en pulgadas o en centímetros. Las variables medidas al nivel de intervalo o de proporción son muy raras en ciencias sociales.

Para terminar: la estructura de una propiedad no tiene que ser totalmente capturada por un instrumento de medición. Tomemos la edad, una propiedad susceptible de ser expresada en proporciones. Puedo medir la edad pidiéndoles a los participantes que indiquen su edad en años, preservando así el nivel de proporción. También les puedo preguntar si están por debajo de los veinte, entre veinte y treinta y nueve, entre cuarenta y cincuenta y nueve o si tienen sesenta o más, asignando los puntajes uno, dos, tres y cuatro. Al crear categorías para la edad ya no sabemos exactamente qué tan viejo es alguien. Podemos decir que las personas en la categoría más alta son más viejas, pero no por cuánto. Categorizando la variable perdimos información a nivel de proporción y de intervalo.

4.04 La medición: tipos de variables

Identificamos los niveles de medición nominal, ordinal, de intervalo y de proporción. Ahora veremos cómo podemos interpretar variables con diferentes niveles de medición. También discutiremos otras maneras de clasificar las variables de acuerdo con sus características de medición.

Las variables **categóricas** diferencian o bien categorías desordenadas o bien categorías ordenadas. Un tipo especial de variable categórica es la variable **binaria** o **dicotómica**. Este tipo de variable tiene exactamente dos categorías, como masculino o femenino, fumador o no fumador, peludo o sin pelo. Las categorías pueden ser naturales, como en la dicotomía masculino / femenino o

creadas por el investigador, como por debajo de 20 años o con 20 años o más. Las variables categóricas con más de dos categorías a veces son llamadas *politómicas*.

En las variables categóricas las diferencias entre valores no son interpretables. Por supuesto las variables nominal y ordinal son ambas categóricas. Entonces ¿cómo podemos interpretar resultados numéricos a partir de variables categóricas? Bueno, supongamos que mido la preferencia por animales en un grupo de mis amigos, asignando los números uno, dos y tres respectivamente a amigos que prefieren perros, gatos o hamsters. No tiene sentido tomar la preferencia media, digamos, 1.2 y decir que mis amigos tienen una baja preferencia por los animales. También podría haber asignado los números en el orden contrario, resultando en una media alta de 2.8. para una variable nominal como la preferencia por los animales, sólo tiene sentido observar las frecuencias, cuánta gente hay en cada categoría.

¿Y qué hay de un test de matemáticas con 10 preguntas que miden la habilidad matemática al nivel ordinal? Supongamos que administro el test a mis amigos y encuentro un puntaje medio de 6.2. ¿Se puede interpretar esta media? Bueno, no sin los puntajes sólo reflejan un orden. Porque puedo reasignar a la persona con el puntaje máximo con al número 15. El orden sigue siendo el mismo de manera que se preservan las relaciones. Por supuesto si hice esto, el puntaje promedio en el test va a ser repentinamente mucho más alto.

El valor es arbitrario y no informativo de las relaciones diferencias entre personas en cuanto a la propiedad de interés. Así que si tenemos una variable ordinal deberíamos apegarnos a las frecuencias y las estadísticas tales como la mediana y la moda.

Las variables categóricas pueden ser contrastadas con variables **cuantitativas**. Las variables cuantitativas nos permiten determinar no sólo que la gente difiere en una cierta propiedad, sino también en qué medida difieren. Las variables de intervalo y de proporción son cuantitativas.

Para variables cuantitativas como la temperatura, el peso y la longitud, no tiene sentido calcular la media y, por ejemplo, comparar las medias de los grupos. Esto, porque la media es influenciada por la distancia entre números. Para variables cuantitativas la distancia entre números realmente corresponde a la distancia entre personas en la propiedad de interés.

Por ejemplo, si yo mido el peso de mis amigos que poseen un gato y el peso de mis amigos que poseen un perro, puedo comprar su peso medio para ver si la gente con gatos es más pesada porque no hacen extra ejercicio sacando a pasear a su mascota.

Una distinción final que debemos ser capaces de hacer es entre las variables **discretas** y **continuas**. Para variables continuas siempre es posible, al menos en teoría, encontrar un valor entre dos valores cualesquiera. Consideremos el peso corporal: si una persona pesa 65 kilogramos y otra pesa 66 kilogramos, podemos fácilmente imaginarnos encontrando a alguien que pese 65.5, o 65.72 o 65.268334. en la medida en que nuestro instrumento de medición sea lo suficientemente preciso, debería ser posible cualquier valor entre el número y el máximo en la escala.

Las variables discretas, por otro lado, sólo pueden tomar un conjunto limitado de valores. Las variables nominales y ordinales son discretas por su naturaleza. Pero las variables cuantitativas

también pueden ser discretas. Tomemos el número de mascotas que alguien ha tenido. Esta es una variable de proporción, porque las diferencias pueden ser comparadas: la diferencia entre dos y cuatro mascotas es la misma que entre una y tres, y porque la proporción puede ser comparada: alguien con cuatro mascotas tiene dos veces más mascotas que alguien con dos.

El conjunto de valores posibles es muy limitado, sin embargo. Empezamos en 0 y después los valores 1, 2, 3, 4. Pero 1.3 mascotas o 4.7 mascotas no son valores válidos. Entonces aquí tenemos un ejemplo de una variable de proporción que es una variable discreta.

La distinción entre variables continuas y discretas es menos relevante, porque no está asociada con cómo podemos interpretar los resultados de la medición, a diferencia de la distinción entre variables categóricas y variables cuantitativas.

4.05 La medición: validez de la medición

Hasta ahora hemos discutido los niveles de operacionalización y medición sin preguntarnos “¿estamos midiendo con precisión, nuestras mediciones reflejan el constructo en que estamos interesados?”. En otras palabras: ¿nuestro instrumento es válido? La validez de un instrumento o método de manipulación comúnmente se llama **validez del constructo** o **validez de la medición**.

¿Cómo evaluamos la validez del constructo? Bueno, supongamos que he creado un cuestionario que busca medir el afecto por los gatos. Un puntaje más alto indica que alguien es más una persona que ama los gatos. ¿Cómo puedo determinar si este puntaje realmente refleja la propiedad “afecto por los gatos”?

Bueno, podemos determinar su **validez superficial**. La validez superficial se refiere a qué tan bien representa el instrumento la propiedad de acuerdo con la evaluación hecha por expertos. Una opción experta puede ser útil en la fase de desarrollo, pero por supuesto los expertos se pueden equivocar. Una alternativa ligeramente mejor es mostrar que el instrumento tiene **validez predictiva** y **validez de criterio**, demostrando que el instrumento puede predecir una propiedad o criterio relevantes. Por supuesto la habilidad de predecir algo no significa que los puntajes usados para la predicción reflejen con precisión el constructo que pretendemos estudiar.

Supongamos que creo un instrumento para medir la motivación que puede predecir satisfacción en el trabajo. Esto no significa que el instrumento mida la motivación, podría reflejar cualquier constructo, por ejemplo, la actitud positiva. De manera que la **validez de criterio** puede apoyar la afirmación de que estamos midiendo *algo* consistentemente, pero tiene un valor limitado a la hora de demostrar que eso que estamos midiendo es de hecho el constructo que queremos medir.

Lo que sería ideal es si ya tenemos un instrumento válido para la propiedad de interés. Podríamos entonces administrar ambos instrumentos y ver si los puntajes en la nueva escala concuerdan con la escala ya validada. Infelizmente, no hay muchos instrumentos patrón para constructos sociales y psicológicos.

Otra solución sería si pudiéramos revisar directamente nuestras mediciones. Consideremos la longitud corporal: podemos usar una cinta métrica y después revisar si la persona cuya cabeza sobresale más tiene la medición más alta. Esta forma puramente cualitativa de evaluar una

propiedad es engorrosa, pero nos permite revisar directamente la validez de la medición a una escala clínica.

Para constructos sociales y psicológicos la situación es muy diferente. No tenemos una manera indisputada y directa de determinar si una persona es más inteligente o afectuosa con los gatos que otra. Entonces, ¿hay otras maneras de evaluar la validez del constructo? Bueno, podemos hacerlo indirectamente, observando si los puntajes se relacionan con variables similares y diferentes de la manera que esperamos. Nos referimos a esto como **validez convergente** y **validez discriminante**.

Por ejemplo, yo puedo esperar puntajes en mi escala de afecto por los gatos que muestren concordancia o convergencia con los puntajes obtenidos de una medición observacional del afecto por los gatos, en la cual la gente pase diez minutos en un cuarto con un gato y contamos el número de veces que la persona mira o acaricia al gato.

Esperaría menos concordancia entre mi cuestionario de afecto por los gatos y un cuestionario de afecto por los animales salvajes. No sería extraño encontrar alguna asociación, pero esperaría que fuera menor. Finalmente, esperaría que no hubiese asociación con una variable que es supuestamente ajena al afecto por los gatos, como el afecto por la pizza.

Un método sistemático para evaluar la validez convergente y la validez discriminante se llama **matriz de métodos y atributos múltiples**. En este abordaje utilizamos diferentes instrumentos, por ejemplo, diferentes cuestionarios u observaciones e instrumentos de autoevaluación para medir dos atributos.

Tomemos el afecto por los gatos y el afecto por la pizza como ejemplos. Esperaríamos una asociación muy alta entre el afecto por los gatos, medido por observación y medido a través de autoevaluación. Y lo mismo para el afecto por la pizza; esperaríamos que los instrumentos de observación y de autoevaluación mostraran una fuerte convergencia.

Esperaríamos una asociación muy baja entre el afecto por la pizza y el afecto por los gatos, ambos medidos utilizando autoevaluación. Una baja asociación es posible porque algunas personas tienden a respuestas socialmente deseables o generalmente positivas en los cuestionarios. La misma asociación muy pequeña puede también ser esperada entre el afecto por los gatos y por la pizza medidos por observación. Finalmente, esperaríamos que no hubiese asociación alguna entre diferentes constructos medidos con diferentes métodos.

Si las relaciones muestran todos los patrones esperados entonces hemos sustentado indirectamente la validez de constructo de esos cuatro instrumentos. Por supuesto este es un proceso complicado, porque la falta de convergencia o de validez discriminante puede deberse a cualquiera de los instrumentos. Esto requeriría un nuevo estudio que combinara los instrumentos con otros instrumentos de maneras nuevas, para averiguar dónde está exactamente el problema. Espero que puedan apreciar cuán desafiante es evaluar la validez de constructo para constructos sociales y psicológicos.

4.06 La medición: confiabilidad de la medición

Un instrumento de medición debería ser válido y también confiable. La **confiabilidad de la medición** se refiere a la **consistencia o estabilidad o precisión** del instrumento. Un instrumento confiable resultará en puntajes altamente semejantes si medimos repetidamente una propiedad estable en la misma persona.

La báscula de mi baño, por ejemplo, no es *perfectamente* confiable. Yo me paro en ella tres veces seguidas y usualmente muestra tres lecturas diferentes. Pero en la medida en que las medidas difieren por cien o doscientos gramos, la confiabilidad de la báscula no es lo suficientemente buena para mí.

Entonces ¿cómo determinamos la confiabilidad de los instrumentos que miden constructos sociales y psicológicos? Bueno, en algunos casos es tan fácil como administrar el instrumento dos veces a un grupo de participantes y determinar qué tan fuertemente coinciden los resultados de la primera y la segunda mediciones. Esto se llama **confiabilidad test-retest**. Podemos usar este método si estamos midiendo cosas como el peso o los tiempos de reacción, pero tan pronto como el recuerdo que tiene una persona de sus respuestas entra en juego, las cosas se ponen más complicadas.

Supongamos que tengo un cuestionario que mide el afecto por los gatos. Consiste en cinco preguntas. Pido un grupo de mis amigos que llenen ese cuestionario una vez y después otra vez quince minutos después. Ellos probablemente recordarán las respuestas que me dieron la primera vez. Ahora, si encuentro una alta consistencia en los puntajes de la primera y la segunda mediciones, ¿esto se debe a que el instrumento es confiable, o a que mis amigos tienen buena memoria y les gusta parecer consistentes en sus actitudes?

Una manera de resolver este problema es mirar la consistencia, no entre diferentes momentos, sino entre diferentes partes del instrumento en una sola ocasión. A esto se le llama **consistencia interna**. Comparamos respuestas en las tres primeras preguntas y en las dos últimas. Por supuesto sólo podemos hacer esto si el instrumento consiste en varias preguntas que se supone que son comparables y miden el mismo constructo.

Si este es el caso podemos determinar la **confiabilidad por mitades divididas**³ dividiendo aleatoriamente en dos y evaluando la asociación entre la primera mitad y la segunda. También hay estadísticas que son equivalentes al promedio de todas las formas posibles de dividir el test.

Si la medición consiste en observación en lugar de autoevaluación, podemos tener dos mediciones del mismo comportamiento y evaluar la asociación entre los dos momentos. A esto se le conoce como **consistencia o confiabilidad intra-observación**.

Por supuesto la memoria del observador puede inflar la asociación. Dado que no importa quién haga las observaciones, se puede también evaluar la confiabilidad de la observación teniendo dos personas diferentes que observan y puntúan el comportamiento y observar la asociación entre los

³ Traducción totalmente libre. La expresión original es “Split-halves reliability” que, traducida literalmente, sería algo así como “confiabilidad por división al medio”.

puntajes de ambos observadores. Llamamos a esto **consistencia o confiabilidad inter-observación**.

Ok. Ya hemos visto diferentes maneras de establecer qué tan confiable o preciso es un instrumento. Pero ¿qué es lo que hace a un instrumento menos confiable? Si el instrumento refleja perfectamente el “**verdadero puntaje**” o verdadero valor en la propiedad de interés, entonces el resultado de la medición o “**puntaje observado**” debería ser el mismo todas las veces.

Pero ¿qué pasa si medimos sistemáticamente un constructo adicional? Tomemos la escala de afecto por los gatos. ¿Qué pasa si estas preguntas también exploran el constructo “actitud positiva general”? esto puede resultar en un puntaje sistemáticamente alto para gente con una actitud positiva. Llamamos a esto **error sistemático**. Esto significa que nuestro instrumento es menos válido, pero no menos confiable.

En la medida en que el puntaje observado es determinado sólo por el “**verdadero puntaje**” en afecto por los gatos y el “**error sistemático**” es causado por el segundo constructo, actitud positiva, entonces todavía podemos obtener el mismo puntaje observado cada vez que medimos a la misma persona.

La confiabilidad es influenciada por el **error aleatorio**, error que es enteramente debido al azar. Si el puntaje observado es en parte determinado por fluctuaciones aleatorias, entonces obtenemos valores diferentes cada vez que medimos a la misma persona.

Si una escala es totalmente no confiable, no hay asociación entre los puntajes observados en diferentes momentos de la medición, por lo tanto, estamos básicamente midiendo error aleatorio o ruido. Puesto de otra manera, esto significa que por lo menos alguna confiabilidad es requerida antes de que un instrumento pueda ser válido. Lo opuesto no es posible. Un instrumento perfectamente confiable puede ser totalmente inválido. Esto pasa cuando el instrumento mide perfectamente un constructo diferente de aquel que se supone que debe medir.

Consideremos las posibilidades en más detalle. Por supuesto que el peor escenario es cuando un instrumento tiene baja confiabilidad y baja validez: un montón de error aleatorio y sistemático. Incluso si el verdadero puntaje contribuye poco al puntaje observado, será casi imposible distinguir esa contribución.

Un instrumento también puede tener baja confiabilidad y alta validez: un montón de error aleatorio pero muy poco error sistemático. Estamos midiendo la propiedad correcta, sólo que de forma muy imprecisa. Un instrumento también puede tener alta confiabilidad y baja validez: una pequeña cantidad de error aleatorio y un montón de error sistemático. Estamos midiendo la propiedad errada de forma muy precisa.

El mejor escenario es la alta confiabilidad y por supuesto la alta validez: una pequeña cantidad de error aleatorio y de error sistemático. El puntaje observado es principalmente determinado por el puntaje verdadero. Estamos midiendo el constructo correcto con gran precisión.

Por supuesto el truco es separar todos esos componentes de error del puntaje verdadero, incluso si hay una cantidad adecuada de error sistemático y error aleatorio. Los expertos en psicometría y

en sociometría buscan hacer esto utilizando modelos estadísticos para poner parcialmente fuera el error aleatorio y el error sistemático.

DOCUMENTO DE TRABAJO