



Universidad Nacional Autónoma de México



Facultad de Psicología

Sobre la Teoría de las Escalas de Medición.

Autor: S. S. Stevens.

Fuente: *Science*. Nueva Serie, Vol. 103, No. 2684 (7 de junio de 1946), pp. 677-680.

Publicado por: Asociación Estadounidense para el Avance de la Ciencia.

Traducido por:

R. Emmanuel Trujano Espinoza & Ramsés Vázquez Lira.

Laboratorio de Comportamiento y Adaptación

(Edificio D, Primer Piso, Cubículo 1).

Facultad de Psicología, UNAM.

Tel. 5622 2281

Sobre la Teoría de las Escalas de Medición

S. S. Stevens

Director del Laboratorio de Psicoacústica, Universidad de Harvard

DURANTE SIETE AÑOS UN COMITÉ de la Asociación Británica para el Avance de la Ciencia debatió acerca del problema de la medición. Citados en 1932 para representar a la Sección A (Ciencias Matemáticas y Físicas) y la Sección J (Psicología), el comité fue requerido para considerar y reportar sobre la posibilidad de "estimadores cuantitativos de eventos sensoriales"-en pocas palabras: ¿Es posible medir la sensaciones humanas? La deliberación llevó sólo a desacuerdos, sobre todo acerca de qué se quiere decir con el término medición. En un reporte interino de 1938 un miembro se quejó de que sus colegas "salieron por esa misma puerta por donde entraron", y en aras de tener otro intento para llegar a un acuerdo, el comité pidió que se continuara por otro año.

Para su reporte final (1940) el comité escogió un elemento común para sus debates, dirigiendo sus argumentos hacia un ejemplo concreto de una escala sensorial. Éste fue la escala *Sone* de sonido (S. S. Stevens y H. Davis, *Hearing*. New York: Wiley, 1938), la cual se usa para medir la magnitud subjetiva de una sensación auditiva con una escala que tiene las propiedades formales de otras escalas básicas, como aquellas usadas para medir longitud y peso. De nuevo los 19 miembros del comité salieron por las mismas rutas que entraron, y sus puntos de vista se situaron entre dos extremos. Un miembro dijo "que cualquier ley que se proponga expresar una relación cuantitativa entre intensidad de una sensación e intensidad de un estímulo no es meramente falsa sino de hecho carente de significado a menos y hasta que se le pueda dar algún significado al concepto de adición aplicado a una sensación" (Reporte Final, pp. 245).

Es claro desde esta y desde otras declaraciones del comité que la cuestión real es el significado de medición. Ésta, para estar seguros, es una cuestión semántica, pero susceptible de discusión sistemática. Tal vez se llegue mejor a un acuerdo si reconocemos que la medición existe en una variedad de formas y que las escalas de medición caen en ciertas clases definidas. Estas clases están determinadas tanto por operaciones empíricas involucradas en el proceso de "medir" como por propiedades formales (matemáticas) de las escalas. Más aún -y esto es de gran importancia para muchas otras ciencias- las manipulaciones estadísticas que legítimamente pueden aplicarse a datos empíricos dependen del tipo de escala en que se clasifican los datos.

CLASIFICACIÓN DE LAS ESCALAS DE MEDICIÓN

Parafraseando a N. R. Campbell (Reporte Final, pp. 340), podemos decir que medición, en sentido amplio, se define como la asignación de numerales a objetos o eventos de acuerdo a ciertas reglas. El hecho de que los numerales puedan ser asignados bajo diferentes reglas nos lleva a diferentes tipos de escalas y, por ende, diferentes tipos de medición. El problema, entonces, se convierte en hacer explícitas (a) varias reglas para la asignación de numerales, (b) las propiedades matemáticas (o estructura de grupo) de las escalas resultantes, y (c) las operaciones estadísticas aplicables a las mediciones hechas con cada tipo de escala.

Las escalas son posibles en primer lugar sólo porque existe un cierto isomorfismo entre qué podemos hacer con los aspectos de los objetos y las propiedades de las series de numerales. Al lidiar con los aspectos de los objetos nos referimos a las operaciones empíricas para la determinación de igualdad (clasificación), para ordenación de rangos, y para determinar cuándo las diferencias y cuándo las razones entre los aspectos de los objetos son iguales. Las series convencionales de numerales nos conducen a operaciones análogas: podemos identificar los miembros de una serie de numerales y podemos clasificarlos. Podemos su orden pues está dado por convención. Podemos determinar diferencias semejantes como $8-6 = 4-2$, y razones semejantes, como $8/4 = 6/3$. El isomorfismo entre estas propiedades de las series de numerales y ciertas operaciones empíricas que realizamos con los objetos permite el uso de las series como un modelo para representar aspectos del mundo empírico. El tipo de escala lograda depende del carácter de las operaciones empíricas básicas realizadas. Estas operaciones están limitadas por la naturaleza de las cosas a representar y por nuestra elección de procedimientos, pero, una vez seleccionadas, las operaciones determinan el uso de una u otra de las escalas listadas en la tabla 11.

¹ Una clasificación esencialmente equivalente a la contenida en esta tabla fue presentada anteriormente en el Congreso Internacional para la Unidad de la Ciencia, Septiembre de 1941. El autor está en deuda con el fallecido Prof. G. D. Birkhoff por la valiosa discusión la cual condujo a la finalización de la tabla en su forma final.

Tabla 1

Escala	Operaciones empíricas básicas	Estructura matemática de grupo	Estadísticos permitidos (invariantes)
NOMINAL	Determinación de igualdad	<i>Grupo de permutación</i> $x' = f(x)$ $f(x)$ significa cualquier sustitución uno-a-uno	Número de casos Modo Correlación de contingencia
ORDINAL	Determinación de mayor o menor	<i>Grupo isotónico</i> $x' = f(x)$ $f(x)$ significa cualquier monotónicamente creciente	Mediana Percentiles
INTERVALAR	Determinación de igualdad de intervalos o diferencias	<i>Grupo lineal general</i> $x' = ax + b$	Media Desviación estándar Correlación de orden de rango Coeficiente de variación
RAZÓN	Determinación de igualdad de razones	<i>Grupo de similitud</i> $x' = ax$	Coeficiente de variación

La decisión de descartar los nombres de escalas comúnmente encontrados en los textos sobre medición está basada en la ambigüedad de términos tales como "intensivos" y "extensivos". Ambas escalas ordinal e intervalar son llamadas en ocasiones intensivas, y por su parte las escalas intervalar y de razón algunas veces son clasificadas como extensivas.

Se notará que la columna que lista las operaciones básicas necesarias para crear cada tipo de escala es acumulativa: Para una operación listada enseguida de una escala en particular se deben agregar todas aquellas operaciones que la preceden. Se puede construir una escala intervalar solo si se prueba que tenemos una operación para determinar igualdad entre intervalos, para determinar más grande o más pequeño que, y para determinar igualdad (ni mayor ni menor). A esas operaciones se les tiene que anexar un método para demostrar igualdad de razones si se quiere construir una escala de razón.

En la columna que registra la estructura de grupo de cada escala se listan las transformaciones matemáticas que permiten conservar invariante la forma de la escala. Entonces cada numeral, x , en una escala puede reemplazarse por otro numeral, x' , donde x' es una función de x listada en esta columna. Cada grupo matemático en la columna está contenido en el grupo que le sigue.

La última columna presenta ejemplos del tipo de operaciones estadísticas apropiado para cada escala. Esta columna es acumulativa en el sentido de que todos los estadísticos listados son admisibles para datos englobados en una escala de razón. El criterio para la pertinencia de un estadístico es invarianza ante las transformaciones en la tercera columna. Entonces el caso de la mediana (punto medio) de una distribución mantiene su posición ante todas las transformaciones que preserven el orden (grupo isotónico), pero

un elemento localizado en la media se mantiene en la media solo ante transformaciones restringidas a aquellas de un grupo lineal. La razón expresada por el coeficiente de variación se mantiene invariante solo ante transformaciones de similitud (multiplicación por una constante). (El coeficiente de correlación de rango de órdenes usualmente se considera apropiado en una escala ordinal, pero de hecho este estadístico asume intervalos semejantes entre rangos sucesivos y por tanto clama por una escala intervalar.)

Ahora bien consideremos a cada una de las escalas.

ESCALA NOMINAL

La *escala nominal* representa la asignación de numerales menos restringida. Los numerales se utilizan solo como etiquetas, y palabras y letras servirían también. En ocasiones se distinguen 2 tipos de asignaciones nominales, como lo ilustran (a) los números de los jugadores de fútbol para identificar a los individuos y (b) la "enumeración" de tipos o clases, donde a cada miembro de una clase se le asigna el mismo numeral. De hecho el primero es un caso especial del segundo, dado que cuando etiquetamos a nuestros jugadores de fútbol nos encontramos con clases de un miembro cada una. Dado que el propósito es alcanzado por igual cuando cualquier par de numerales son intercambiados, la forma de esta escala permanece invariante bajo la sustitución general o el grupo de permutaciones (en ocasiones llamado el grupo simétrico de transformaciones). El único estadístico relevante en escalas nominales de tipo A es el número de casos, por ejemplo la cantidad de numerales asignados a los jugadores. Pero una vez que se han formado clases que contienen varios individuos (tipo B), se puede determinar la clase más numerosa (el modo), y bajo ciertas condi-

ciones se puede probar con métodos de contingencia hipótesis referentes a la distribución de casos entre las clases.

La escala nominal es una forma primitiva, y naturalmente hay muchos que reclamarán que es absurdo atribuir a este proceso de asignar numerales la dignidad implicada por el término medición. Ciertamente no puede haber desacuerdo con esta objeción, debido a que nombrar cosas es una cuestión arbitraria. Sin embargo le damos nombre, el uso de numerales como nombres de clases es un ejemplo de "asignación de numerales de acuerdo a una regla". La regla es: No asignar el mismo numeral a diferentes clases o diferentes numerales a la misma clase. Fuera de eso, cualquier cosa es válida en la escala nominal.

ESCALA ORDINAL

La *escala ordinal* surge de la operación de ordenación de rangos. Dado que cualquier transformación de "preservación de orden" dejaría la forma de la escala invariante, esta escala tiene la estructura de lo que podemos llamar el grupo de preservación de orden ó isotónico. Un clásico ejemplo de una escala ordinal es la escala de dureza de los minerales. Otros ejemplos se encuentran dentro de las escalas de inteligencia, personalidad, el grado o calidad de una tela, etc.

El hecho en cuestión, es que la mayoría de las escalas empleadas amplia y efectivamente por psicólogos son las escalas ordinales. En estricto sentido los estadísticos comunes como la media y la desviación estándar no deben de ser usadas con estas escalas, debido a que estos estadísticos implican un conocimiento de algo más que el orden de rangos relativos de los datos. Por otro lado, para esta matematización "ilegal" puede mencionarse un tipo de sanción pragmática: En numerosas ocasiones nos lleva a resultados fructíferos. Mientras que la restricción de este procedimiento no sirva a propósito alguno, es apropiado puntualizar que la media y la desviación estándar computadas en la escala ordinal son un amplio error en respecto a que los intervalos sucesivos en la escala son desiguales en tamaño. Cuando sólo se conoce el orden del rango de los datos, hay que proceder con cautela con nuestros estadísticos, y en especial con las conclusiones que extraemos de ellos.

Incluso cuando se aplican aquellos estadísticos que normalmente son apropiados para escalas ordinales, a veces encontramos el rigor comprometido. Así pues, aunque se indica en la Tabla 1 que las medidas percentiles pueden aplicarse para datos ordenados por rangos, cabe señalar que el procedimiento habitual de asignar un valor a un percentil por medio de interpolación lineal en un intervalo de clase está, en el estricto sentido, totalmente fuera de los límites. Del mismo modo, no es estrictamente apropiado determinar el punto medio de un intervalo de clase por interpolación lineal, ya que la linealidad de una escala ordinal es precisamente la propiedad que está abierta a controversia.

ESCALA INTERVALAR

Con la *escala intervalar* llegamos a la forma "cuantitativa" en el sentido habitual de la palabra. Casi todas las mediciones estadísticas usuales son aplicables, a menos que sean del tipo que implican conocimiento de un punto cero "verdadero". El punto cero en una escala intervalar es cuestión de convención o conveniencia, como lo demuestra el hecho de que la forma de la escala permanece invariante cuando una constante es agregada.

Este punto es ilustrado por nuestras dos escalas de temperatura, la Celsius y la Fahrenheit. Intervalos de temperatura semejantes están fuera de escala observando volúmenes iguales de expansión; un cero arbitrario es agregado sobre cada escala; y un valor numérico sobre una de las escalas es transformado a un valor en la otra por medio de una ecuación de la forma $x' = ax + b$. Nuestras escalas de tiempo ofrecen un ejemplo similar. Fechas en un calendario son transformadas por esta misma ecuación de una forma semejante. En estas escalas, por supuesto, no tiene sentido el decir que un valor es el doble, o alguna otra proporción mayor a otro valor.

Los *periodos de tiempo*, sin embargo, pueden ser medidos en escalas de razón y un periodo puede ser definido correctamente como el doble de otro. Lo mismo es probablemente cierto para medidas de temperatura cuantificadas en la bien conocida Escala Absoluta.

La mayoría de las mediciones psicológicas aspiran a crear escalas intervalares, y en algunas ocasiones lo logran. El problema es idear operaciones para igualar las unidades de las escalas —un problema no siempre fácil de resolver, pero uno para el cual existen algunos posibles modos de atacarlo. Solo ocasionalmente existe consternación por la localización de un punto cero "verdadero", debido a que los atributos humanos medidos por psicólogos usualmente existen en un grado positivo, el cual es grande comparado con el rango de su variación. Respecto a esto, los atributos son análogos a la temperatura siendo que se encuentra en la vida diaria. La inteligencia, por ejemplo, es útilmente evaluada en escalas ordinales, las cuales intentan aproximarse a escalas intervalares, y no es necesario definir que significa un puntaje cero de inteligencia.

ESCALA DE RAZÓN

Las *escalas de razón* son aquellas ampliamente encontradas en física y sólo son posibles cuando existen operaciones para determinar todas las cuatro relaciones: igualdad, ordenación de rangos, igualdad de intervalos e igualdad de razones. Una vez que la escala es erigida, sus valores numéricos pueden ser transformados (como pulgadas a centímetros) sólo al multiplicar cada valor por una constante. Un cero absoluto siempre está involucrado, incluso aunque el valor cero en algunas escalas (por ejemplo, temperatura absoluta) puede que nunca se genere. Todos los tipos de mediciones estadísticas son aplicables a las escalas de razón, y sólo con estas escalas podemos propiamente satisfacer transformaciones logarítmicas como aquellas involucradas en el uso de decibeles.

Principalmente, entre las escalas de razón, es la escala de números en sí —números cardinales— la escala que utilizamos cuando contamos cosas como monedas, manzanas, fichas. Esta escala de la numerosidad de agregados es tan básica y tan común que normalmente ni se menciona en discusiones sobre medición.

Es convencional en física el distinguir entre dos tipos de escalas de razón: *fundamentales* y *derivadas*. Las escalas fundamentales están representadas por la longitud, el peso y la resistencia eléctrica, en tanto que las escalas derivadas están representadas por la densidad, la fuerza y la elasticidad.

Estás últimas son magnitudes *derivadas* en el sentido de que son funciones matemáticas de ciertas magnitudes fundamentales. De hecho, en física estas mediciones se han vuelto más numerosas que las mediciones fundamentales, las cuales comúnmente se han mantenido como básicas porque satisfacen el criterio de *aditividad*. Pesos, longitudes y resistencia pueden ser sumadas en el sentido físico, pero este hecho empírico es generalmente mantenido con gran prominencia en teoría de medición, más que lo que merece en realidad. Las llamadas escalas fundamentales son ejemplos importantes de escalas de razón, pero solo son ejemplos. De hecho, se puede demostrar que las escalas fundamentales pueden ser creadas incluso si las operaciones físicas de adición son descartadas como un desempeño imposible. Dadas tres balanzas, por ejemplo, cada una teniendo una apropiada construcción, un conjunto de pesos estándar puede ser manufacturado sin que si acaso sea necesario poner dos pesos en la misma escala al mismo tiempo. El procedimiento es muy extenso para describirlo en estas líneas, pero su factibilidad se menciona aquí simplemente para sugerir que la adición física, aunque algunas veces es posible, no necesariamente es la base de toda medición.

Escalas de razón de magnitudes psicológicas son poco comunes pero no totalmente desconocidas. La escala *Sone* discutida por el Comité Británico es un ejemplo fundado en un intento deliberado de que observadores humanos juzgaran las razones auditivas de pares de tonos. El juicio de intervalos semejantes ha sido establecido como un método legítimo, y con el trabajo sobre razones sensoriales, el cual comenzó en varios laboratorios independientes, el paso final fue el tomar la asignación de numerales a sensaciones de sonido de tal manera que las relaciones entre las sensaciones se reflejan en las relaciones aritméticas ordinarias en las series de numerales. Como en toda medición existen límites impuestos por el error y la variabilidad, pero dentro de estos límites la escala *Sone* tiene que estar clasificada apropiadamente como una escala de razón.

Para el comité Británico, entonces, podemos arriesgarnos a sugerir a manera de conclusión que la definición más útil y liberal de medición es, como uno de sus miembros recomienda, "la asignación de numerales a cosas para representar hechos y convenciones acerca de ellos". El problema sobre qué es y qué no es medición entonces se reduce a una simple pregunta: ¿Cuáles con las reglas, si las hay, bajo las cuales se

asignan los numerales? Si podemos señalar un conjunto consistente de reglas, obviamente estamos interesados en medición de algún tipo, y por tanto podemos proceder a cuestiones más interesantes acerca de qué tipo de medición es. En la mayoría de los casos la formulación de las reglas de asignación revela directamente el tipo de medición y por lo tanto el tipo de escala involucrados. Si cualquier ambigüedad permanece, podemos buscar la respuesta final y definitiva en la estructura de grupo matemática de la forma de la escala: ¿De que maneras podemos transformar sus valores y aún así preservar todas las funciones previamente cumplidas? Sabemos que los valores de todas las escalas pueden ser multiplicados por una constante, lo cual cambia el tamaño de la unidad. Si además una constante puede ser sumada (o elegir un nuevo punto cero), es prueba positiva que no estamos ante una escala de razón. Entonces, si el propósito de la escala todavía se cumple cuando sus valores son elevados al cuadrado o al cubo, esto ni siquiera es una escala intervalar. Y finalmente, si dos valores cualesquiera pueden ser intercambiados a voluntad, la escala ordinal se descarta y la escala nominal es la única posibilidad restante.

Esta solución propuesta al problema semántico no es dar a entender que todas las escalas pertenecientes al mismo grupo matemático son igualmente precisas ó certeras ó útiles ó "fundamentales". La medición nunca es mejor que las operaciones empíricas por las cuales se lleva a cabo, y las operaciones van de malas a buenas. Se puede objetar cualquier escala particular, sensorial o física, con argumentos sobre sesgo, baja precisión, generalidad restringida, y otros factores, pero el objetor debería recordar que éstas son cuestiones relativas y prácticas, y que ninguna escala usada por mortales está perfectamente libre de error.

