Scales of Measurement

Nathaniel E. Helwig

Associate Professor of Psychology and Statistics
University of Minnesota

August 29, 2020

Table of Contents

Table of Contents

Importance of Measurement

What it means to "measure" something has long been a topic of both scientific and philosophical debate.

The concept of measurement is fundamental to the field of psychology because we need reliable measurements of psychological constructs in order to trust any statistical results pertaining to those constructs.

Statistical methods cannot overcome issues pertaining to poor measurement (garbage in, garbage out principle).

Stevens' (1946) Definition of Measurement

In the influential paper "Theory of Scales of Measurement", Stevens (1946) defined measurement as "the assignment of numerals to objects or events according to rules" (p. 677).

This broad (yet controversial) definition is used in many psychological studies, as well as many fields throughout the social sciences.

- See Michell (1986) for a critique of this definition

In his paper, Stevens presents four different scales (or levels) of measurement that can characterize different types of measures that are used in psychological and other social science studies.

Table of Contents

Nominal Measurement Scale

According to Stevens (1946), "[t]he *nominal scale* represents the most unrestricted assignment of numerals" such that "[t]he numerals are used only as labels or type numbers, and words or letters would serve as well" (p. 678).

Nominal scales of measurement involve assigning numerals that are *not* meant to convey any quantitative meaning.

Example. Suppose that we record the variable Gender, and code the responses as 1 = Female, 2 = Male, and 3 = Other.

Variables that are measured using a nominal scale are discrete categorical variables that have probability mass functions.

Ordinal Measurement Scale

According to Stevens (1946), "[t]he *ordinal scale* arises from the operation of rank-ordering" such that "any 'order-preserving' transformation will leave the scale form invariant" (p. 679).

Ordinal scales of measurement involve assigning numerals that are only meant to convey meaning regarding the order of objects or events.

Example. The positions in which runners cross the finish line for a race, i.e., first place, second place, third place, etc.

Variables that are measured using an ordinal scale are discrete (ordered) categorical variables that have probability mass functions.

Interval Measurement Scale

According to Stevens (1946), "[w]ith the *interval scale* we come to a form that is "quantitative" in the ordinary sense of the word." (p. 679).

Interval scales are what we typically think of when we think of a quantitative measure, but such scales have a zero point that is "a matter of convention or convenience" (Stevens, 1946, p. 679).

Example. Celsius and Fahrenheit scales used to measure temperature.

- $^\circ\text{Fahrenheit} = {}^\circ\text{Celsius}\left(\frac{9}{5}\right) + 32$

Variables that are measured using an interval scale are continuous variables that have probability density functions.

Ratio Measurement Scale

According to Stevens (1946), "*ratio scales* are those most commonly encountered in physics and are possible only when there exist operations for determining all four relations: equality, rank-order, equality of intervals, and equality of ratios" (p. 679).

Like interval scales, except that ratio scales have a true zero point, i.e., a value of 0 indicates an absence of the property being measured.

Example. Consider the measurements of length and weight.

- 1 foot = 12 inches (0 inches means lack of length)
- 1 pound = 16 ounces (0 oucnces means lack of weight)

Scales of Measurement Summary

Table 1: Reproduction of Table 1 from Stevens (1946).

Scale	Basic Empirical Operations	Mathematical Group Structure	Permissible Statistics (invariantive)
NOMINAL	Determination of equality	*Permutation group* $x' = f(x)$ $f(x)$ means any one-to-one substitution	Number of cases Mode Contingency correlation
ORDINAL	Determination of greater or less	*Isotonic group* $x' = f(x)$ $f(x)$ means any monotonic increasing function	Median Percentiles
INTERVAL	Determination of equality of intervals or differences	*General linear group* $x' = ax + b$	Mean Standard deviation Rank-order correlation Product-moment correlation
RATIO	Determination of equality of ratios	*Similarity group* $x' = ax$	Coefficient of variation

Note. According to Stevens (1946) "any numeral, x, on a scale can be replaced by another numeral, x', where x' is the function of x listed in this column" (p. 678).

Table of Contents

The Quality of a Measurement

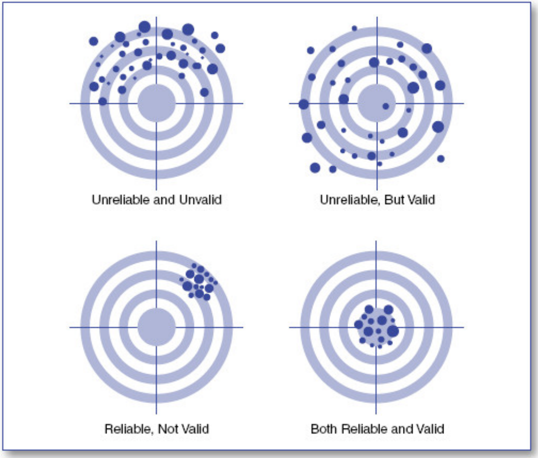

Figure 1: Visualization of reliability and validity from Ruel et al. (2016).

Definition of Reliability

A *reliable* measure is one that is "dependable, replicable, and consistent" (Ruel et al., 2016).

A reliable measure is one that produces the same measurement results (up to the scale's precision) when measuring two objects that have the same amount of the property being measured.

Example. If two individuals have the same weight, a reliable scale would return the same weight measurements up to the scale's measurement precision (e.g., 0.1 pounds).

Types of Reliability

- *Test-retest reliability*: the correlation between two measurements of the same object measured at different times using the same scale.
- *Alternate form reliability*: the correlation between two measurements of the same object measured at the same time using different scales.
- *Internal consistency*: the pairwise correlations between the individual items that compose the measurement scale (item-wise congruence).
- *Split-test reliability*: the correlation between the scores on the first half and the second half of the measurement scale.
- *Inter-rater reliability*: the correlation between measurements as determined by two independent subjects (raters) measuring the same object.

Definition of Validity

A *valid* measure is one that "operates the way [researchers] expect" (Ruel et al., 2016).

A valid measure is one that measures what it is supposed to measure without missing key properties or including unintended properties.

Example. If an exam is supposed to measure statistical knowledge, then the exam would be a valid measurement if it comprehensively quantifies statistical knowledge without measuring extra unintended constructs (e.g., reading or language skills).

Types of Validity

- *Face validity*: the measurement appears valid at face value.
- *Content validity*: the content of the measurement scale is complete, applicable, and representative of the measured construct.
- *Criterion-based validity*: the agreement between a scale's measurement and the measurement from a "gold standard" scale.
- *Concurrent validity*: the agreement between a scale's measurement and measurements of related (distinct) constructs measured from the same objects.
- *Predictive validity*: the ability of a measurement to predict related constructs.
- *Construct validity*: the degree to which a measurement scale is assessing the construct of interest, e.g., instead of some other construct.
- *Convergent validity*: the agreement between two measures in the same study that are intended to assess the same construct
- *Discriminant validity*: the lack of agreement between two measures in the same study that are intended to assess different constructs

References

Michell, J. (1986). Measurement scales and statistics: A clash of paradigms. *Psychological Bulletin 100*(3), 398–407.

Ruel, E., W. Wagner III, and B. Gillespie (2016). The quality of measurement: reliability and validity. In E. Ruel, W. Wagner III, and B. Gillespie (Eds.), *The Practice of Survey Research: Theory and Applications*, pp. 78–100. SAGE.

Stevens, S. S. (1946). On the theory of scales of measurement. *Science 103*(2684), 677–680.