PEP 6305
Measurement in Health and Physical Education

Scale Types

Scaling is the process of assigning numerical values, or *scores*, to observations. A *scale* is reference range of values that has a *unit of measurement* appropriate to the construct being measured (McDonald, 1999). A unit of measurement gives meaning to the numerical values, which allows for interpretation of scores.

Scale Types

S.S. Stevens, a Harvard psychologist, published a classic paper in *Science* (1946) in which he described the properties of four scale types: nominal, ordinal, interval, and ratio. These scale types, described in more detail below, are defined by (1) their relation to the properties of the real numbers and (2) how they are affected by transformations (mathematical operations).

The properties of the real numbers are order, magnitude, and origin (i.e., a true zero value). Invariance to transformation means that scale structure and score meaning is maintained despite performing certain mathematical functions on the scores.

Transformation is important when using scores in statistical analysis. Statistical tests, including those used in measurement methods, have certain assumptions about the data, and the interpretation of the results of the statistical test depends on the type of scale to which the measures belong.

The scale type convention is artificial because we are "mapping" observations to the real numbers. The real numbers are only a symbol of what we have observed; they are not the phenomena or behaviors observed. The key is to determine how much these numbers represent the real-world observations, particularly when those numbers are manipulated using mathematical formulas.

Much real data does not fit neatly into one scale type category. The system is useful, however, in (1) relating observations to the set of real numbers and (2) describing the type of information that data may contain. Both of these issues are foundational in measurement, and both issues are critical to interpreting statistical results properly.

Nominal Scales

Nominal scales use numbers as labels, for example: 1, 2, and 3 may be used to label three different groups of subjects. Nominal scales may also use symbols other than numbers (A, B, C; insect, reptile, mammal; etc.) without a loss of information.

The assigning of symbols is mutually exclusive (each case is assigned to one and only one category) and exhaustive (every case is in a category). Each unique category can be assigned a unique number, as in the first example above, but those numbers have no mathematical meaning. The scores are *categorical*, meaning that they can only assume specific discrete integer values, rather than *continuous*, meaning scores that can assume any of the infinite real number values (e.g., 1.3, 2.95, 3.401, etc.).

Nominal-scale scores have none of the properties of the real numbers. Nominal scales are not ordered; in fact, the categories can be re-labeled in a different order and their meaning will not change. The magnitude (or distance) between nominal score values has no meaning; a person in Group 2 does not necessarily possess twice as much of some trait as a person in Group 1. Last, nominal scores have no origin; a score of 0 does not necessarily imply an absence of the construct.

Any transformation or mathematical function that allows a one-to-one substitution of numbers, or any other symbol, regardless of whether the function reorders the symbols or not, preserves the meaning of nominal-level scores. This transformation rule is the least restrictive among the four types of scales. A nominal scale is the least informative from a measurement perspective because it gives no indication of relative magnitude of differences among the scores.

Many mathematical functions make one-to-one substitutions and thereby preserve the meaning of transformed nominal scores. A function, f, performed on a variable, X, produces a one-to-one transformation if

$$f(X_1) = f(X_2) \text{ iff (if and only if) } X_1 = X_2 \tag{1}$$

where $f(X) \neq c$ (c being a constant value). Equation 1 represents one-to-one transformation because

$$\text{if } f(X_1) = f(X_2) \text{ but } X_1 \neq X_2,$$

then f would be a rule that can assign the same transformed value, $f(X_1) = f(X_2) = Y_1$, to more than one score, X_1, X_2, etc., which means it is not a one-to-one transformation. Similarly, if

$$f(X_1) \neq f(X_2) \text{ but } X_1 = X_2,$$

then f is a rule that assigns different transformed values, $f(X_1)$ and $f(X_2)$, to the same score value, $X_1 = X_2 = X$, which violates the definition of a one-to-one transformation.

A permissible transformation function, f, can thus only be any rule that maps unique values in one set of scores to unique values in the second set of transformed scores. For example, a rule which simply re-labels the categories is such a function, so long as the re-labeling rule is mutually exclusive and exhaustive (e.g., [1, 3, 9] may become [1, 9, 3], or [3, 9, 1], or [42, 16, π], or [A, B, C]).

Note that transforming nominal scores does not produce scores that have real number properties with respect to a score's meaning. Transforming nominal scores is equivalent to changing the labels "A, B, C" to "D, E, F": the symbol (the label) has changed, but the meaning (the category) has not changed. This property of nominal scores means that certain statistical calculations and tests may yield numerical results that are not interpretable (Maxwell & Delaney, 1985). For example, if 10 women in a sample are labeled "1" and 10 men in the sample are labeled "2", one could report the "mean gender" as "1.5", but that statement has no meaning.

Nominal scales are common in exercise science. For instance, a goal of many studies is to find a good way to classify people into groups based on certain variables. The possible group classifications, or categories, are nominal-scale scores. Similarly, subjects may be grouped into categories, such as gender or grade level, to determine if persons in those categories differ with respect to some other property or trait.

Ordinal Scales

Ordinal scales have scores that reflect order in meaning; a higher value score means more of the property is present. The concept of order is powerful, and ordinal-scale measures can yield valuable information.

The magnitude of the distance between ordinal-scale scores, however, has no meaning. Also, an ordinal scale does not necessarily have an origin; a score of 0 does not necessarily mean an absence of the property. Like nominal scores, ordinal scores are categorical, which affects the selection and interpretation of statistical tests.

Any order-preserving transformation will maintain the meaning of ordinal scores. Order-preserving functions are *monotonic*, meaning that the function either constantly increases or constantly decreases as the values of its argument increases; the relative order of the scores is not affected (Figure 1).

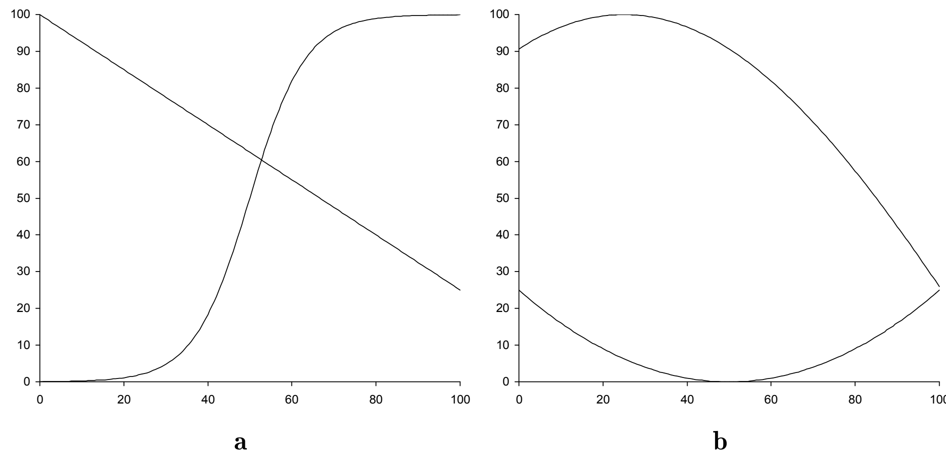

Figure 1 Monotonic (a) and nonmonotonic (b) transformations.

Thus, in addition to meeting the requirements in Equation 1.1 (one-to-one substitution), a monotonic transformation f for ordinal scores must also satisfy

$$f(X_1) > f(X_2) \text{ iff } X_1 > X_2 . \tag{2}$$

Any transformation function must preserve the order of the scores; if X_1 is larger than X_2 before the transformation, the respective transformed scores must maintain that relation.

Ordinal scores are generated in a number of common ways. When the score is a result of ranking subjects, such as occurs in race results (first place, second place, third place, etc.), the scale-level is ordinal. A person who finishes in fourth place is not half as fast as a person who finishes in second place, and there is no "zeroth" place.

Ordinal-level scores also result from (1) a set of judges ranking the performance of many subjects, (2) subjects or judges using Likert-type ratings (e.g., Strongly Agree, Agree, No Opinion, Disagree, Strongly Disagree, etc.), (3) the investigator's categorization of observations that are interval or ratio in nature, and (4) voting. None of these methods yields any information regarding the magnitude of the differences among the various ratings. Ordinal scores only yield information regarding order—such as which scores are "larger" and which are "smaller."

Interval Scales

Interval scales have scores that are ordered and have the additional property that the magnitude of the distance between scores is meaningful. For example, the meaning of the relative difference between scores of "3" and "4" is the same as the meaning of the relative relative difference between scores of "8" and "9"; the unit of measurement itself has interpretable meaning. Interval scores are also continuous rather than categorical, which means that the magnitudes of partial distances (½, ¼, etc.) between scores are also meaningful. An interval scale, however, does not necessarily have an origin; as in nominal and ordinal scales, a score of 0 does not necessarily imply an absence of the property.

Generally, interval scores maintain meaning under transformations using functions, let's call them f and g , that, in addition to the properties in Equations 1.1 and 1.2, preserve the relative magnitudes between two scores:

$$f(X_1) - f(X_2) = b\left[g(X_1) - g(X_2)\right] = c\left(X_1 - X_2\right), \tag{3}$$

where b and c are called *scaling factors*. Scaling factors are multiplicative constants that change the initial unit of measurement into a different unit of measurement, for example, from centimeters to inches. The computed differences in the magnitudes between two scores depends only on the chosen scale—that is, it depends on the unit of measurement. The relative magnitude between scores, and therfore the meaning between scores, is invariant.

Most commonly, interval scores are transformed using a *linear equation*,

$$f(X) = bX + a \quad (b \neq 0) \tag{4}$$

where X is the original interval score, b is the slope ("steepness" of the line's incline), a is the intercept (where the line crosses the Y-axis), and $f(X)$ is the linearly transformed score, which is also an interval score.

The values of b and a change the unit of measurement and location, respectively, of a score. The unit of measurement is what defines a scale; for this reason, b is the scaling factor that changes the *scale* of X. If b is greater than 1, the unit of measurement for $f(X)$ is larger than the unit of measurement for x; if b is less than 1, the $f(X)$-unit is smaller; if b is equal to 1, the units for the $f(X)$-scale and X-scale are equal (Figure 2).

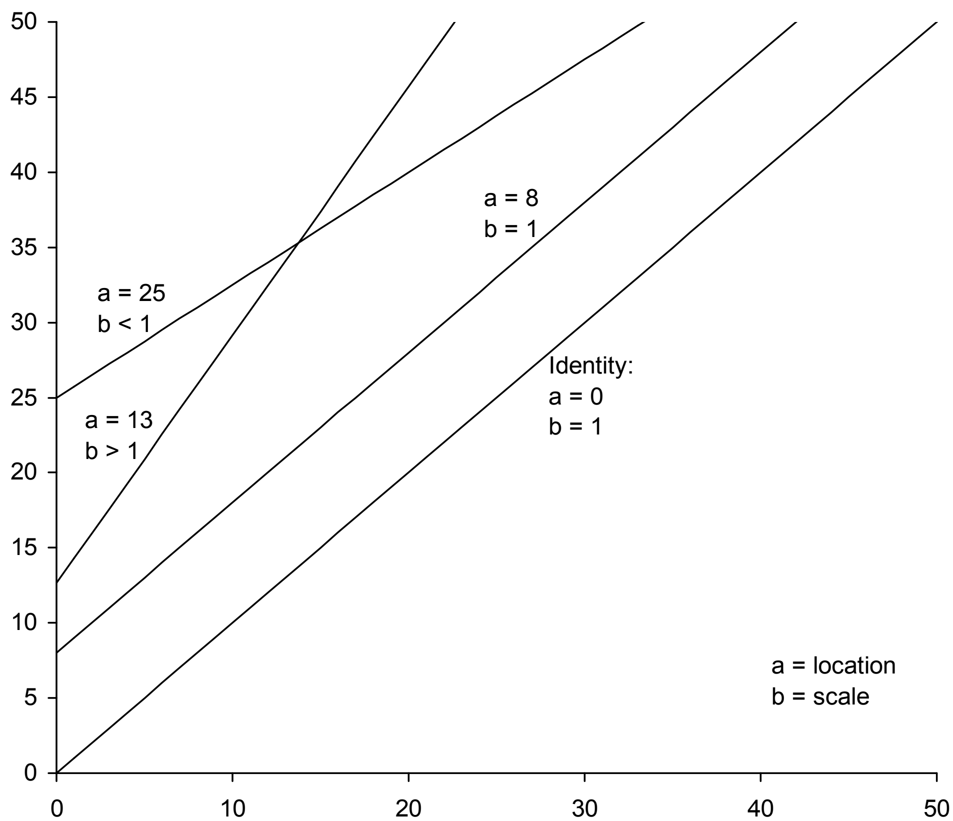

Figure 2 Linear transformations of scale and location.

The term *location* refers to where the line that is the plot of $f(X)$ as a function of X crosses the Y-axis (Figure 2). If a is 0, the line goes through the origin (Cartesian point [0,0]). Any other value of a is the value of $f(X)$ for an X of 0. Linear transformations are used to convert traditional system units into metric system units (SI units), and vice-versa, because both systems yield interval-level measures.

For example, the Celsius temperature scale is an interval scale. The difference between a measure of 50 degrees and a measure of 100 degrees is the same as the difference between a measure of 200 degrees and a measure of 250 degrees: the difference is 50 degrees. Note that when expressed in degrees Fahrenheit, also an interval scale, the magnitudes of the differences for those scores are also equivalent to one another (212 – 122 = 90; 482 – 392 = 90). The Fahrenheit magnitudes differ from the Celsius magnitudes only by a scaling factor (5/9). The relative magnitude of differences between levels is the same. And in either scale a measure of 0 degrees does not imply a complete absence of heat.

Ratio Scales

Ratio scales have all of the properties of the real numbers: order, magnitude, and, unlike the other three scale types, an origin. Ratio scores are also, like the real numbers, continuous, meaning they have an infinite range.

Functions that preserve the meaning of ratio scores meet the properties in Equations 1, 2, and 3, and have the added property:

$$f(X_i) = 0 \text{ for all } X_i = 0 \,, \tag{5}$$

meaning that the transformation preserves the origin of the original scale, the "true zero" score.

Because the scale has a meaningful origin, only multiplication and division (i.e., changes in the unit of measurement via a scaling factor) will maintain the structure of the scale and meaning of the scores:

$$Y = bX \quad (b \neq 0)\,. \tag{6}$$

Equation 6 is Equation 4 with a = 0: a has to equal zero, or the meaning of the origin ("zero point") would be lost. Ratio scores have meaning attached to their order, magnitude, and origin; these features also represent the property that they purport to measure. Rescaling the unit of measurement does not affect these features.

As in Equation 4, the parameter b in Equation 6 changes the unit of measurement, or the scale, of X. A b greater than 1 expands the unit of measurement, a b less than 1 compresses it, and a b equal to 1 has no effect (Figure 3).

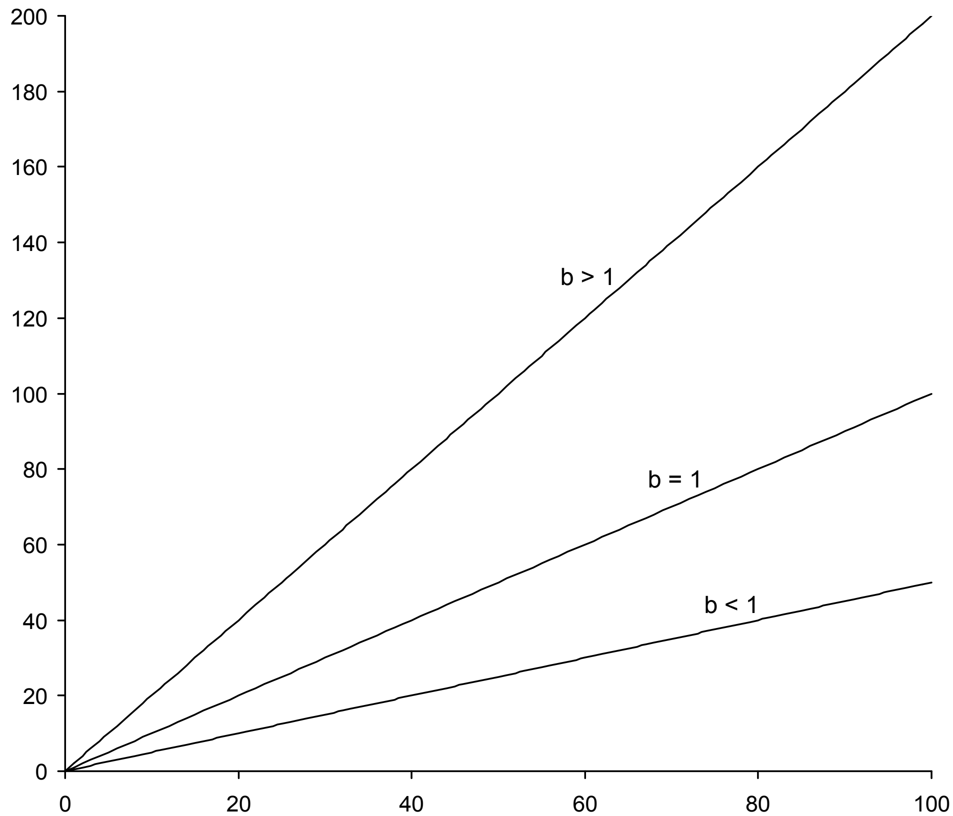

Figure 3 Scalar transformations.

Summary of Scale Types

The relations of the properties among the four scale types can be demonstrated as:

$$\left.\begin{array}{l}\left.\begin{array}{l}\left.\begin{array}{l}\left.f(X_1) = f(X_2) \text{ iff } X_1 = X_2 \right\}\text{nominal} \\ f(X_1) > f(X_2) \text{ iff } X_1 > X_2\end{array}\right\}\text{ordinal} \\ f(X_1) - f(X_2) = b\left[g(X_1) - g(X_2)\right] = c\left(X_1 - X_2\right)\end{array}\right\}\text{interval} \\ f(X_i) = 0 \text{ for all } X_i = 0\end{array}\right\}\text{ratio} .$$

Each successive type scale incorporates the features of the scale types before it.

Nominal and ordinal scores are *discrete*, typically taking only integer values. Interval and ratio scores are *continuous*, assuming any real number value.

Stevens (1946) matched certain statistical methods to data of certain scale types; this point has been argued ever since (Boneau, 1961; Burke, 1953; Gaito, 1980; Michell, 1986). Without getting into the protracted technical arguments, the basic concept is that the interpretation of statistical analyses cannot exceed the complexity of the data. One cannot achieve more detail after analysis than one has before; statistics reduces data, it does not expand it. Information is always lost.

Thus, if the data are nominal or ordinal, results can't be interpreted as intervals or ratios (such as "the average gender was 1.5 ± 0.7"). Scale type does not prohibit performing any particular type of statistical analysis, because numbers are numbers, but scale type is important to interpreting the results of those analyses (Lord, 1953).

Exercises

Exercise 1 Give the construct being measured and the scale type, and defend your answer for the following: (a) number correct on a multiple-choice knowledge test; (b) weighted sum for an attitude survey; (c) scores on the GRE; (d) place in an poetry contest; (e) jersey number of a basketball player; (f) body weight.

Exercise 2 For each of the four scale types, give two examples of measures that are used in your field of study.

References

Boneau, C. A. (1961). A note on measurement scales and statistical tests. *American Psychologist, 16*, 260-261.

Burke, C. J. (1953). Additive scales and statistics. *Psychological Review, 60*(1), 73-75.

Gaito, J. (1980). Measurement scales and statistics: resurgence of an old misconception. *Psychological Bulletin, 87*(3), 564-567.

Lord, F. M. (1953). On the statistical treatment of football numbers. *American Psychologist, 8*, 750-751.

Maxwell, S. E., & Delaney, H. D. (1985). Measurement and statistics: an examination of construct validity. *Psychological Bulletin, 97*(1), 85-93.

McDonald, R. P. (1999). *Test Theory: A Unified Treatment*. Mahwah, NJ: Lawrence Erlbaum Associates.

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

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