SAGE reference

Encyclopedia of Research Design

Variable

Contributors: J. Scott Brown
Edited by: Neil J. Salkind
Book Title: Encyclopedia of Research Design
Chapter Title: "Variable"
Pub. Date: 2010
Access Date: February 7, 2017
Publishing Company: SAGE Publications, Inc.
City: Thousand Oaks
Print ISBN: 9781412961271
Online ISBN: 9781412961288
DOI: http://dx.doi.org/10.4135/9781412961288.n490
Print pages: 1604-1606

Simply put, a variable is a measurement of something that holds at least two distinct values across participants within a study. In contrast, a constant holds the same value across all study participants. Whereas constants such as the speed of light are frequently important for analyses in the natural sciences, the focus of social and behavioral sciences rarely concerns itself with constants. Thus, variables are the basic currency of behavioral research.

The Nature of Variables

Any discussion of variables necessarily must focus on two distinct aspects of these measures: (1) the attributes of variables and how they are measured and (2) the use of variables in scientific analyses. The former refers to the specifics of how the variation of a measure can be described and how variables differ based on their "level of measurement." The latter refers to the utilization of variables in both research design and statistical analysis.

Variable Attributes

All variables must include at least two distinct values that differ across research subjects. These values, or categories, are the characteristics that describe the item of interest. For example, a fundamental component of the basic experimental design is the treatment variable. This measure consists of at least two possible categories: "treatment" and "no treatment" (or "control"). A more complex experimental design might also include a "placebo" group as a third value. Quite commonly, variables can include substantially larger numbers of values ranging from modestly more complex Likert scaling, which typically uses a five-category coding from "strongly disagree" to "strongly agree," to extremely detailed measures such as IQ or income. These more detailed measures are frequently encountered when experimental design is not feasible and when researchers use some form of survey research.

Regardless of how detailed and complex the attributes of any variable becomes, all variables are subject to two requirements. First, the values for any variable must be mutually exclusive from one another. In other words, each subject can have only a single value on each variable in the study. For example, it would be inappropriate for the values of a variable to be male, female, and Latino because it is possible (and, in this case, quite likely) that many individuals might be both male *and* Latino or female *and* Latino. On the other hand, a variable that includes the more logical grouping of only male and female categories would satisfy the mutual exclusivity requirement because one cannot simultaneously be a male and a female.

Second, the attributes of any variable must be exhaustive. This requirement does not mean that every possible value of a variable must be included in the measure. Rather, this requirement is satisfied if all research participants can be assigned, or self-selected, into a value provided. For example, a variable that measures a subject's race might include values such as "white," "black," "Asian," and "Native American/American Indian." The categories in this case would not be exhaustive because some individuals, such as a person of Australian Aboriginal descent or a multiracial person, would not fit into any single category. A simple fix in this particular case would be to add a "multiracial" category as well as an "other" category. The inclusion of residual categories like "other" in the preceding example is commonplace when such values represent extremely small portions of the population being studied or when the actual values of the residual groups are not directly relevant to the research question at hand.

Levels of Measurement

Although only the two requirements of variable attributes noted previously define what a variable must be, several other aspects of variable attributes serve to distinguish types of variables. These are usually referred to as the four levels of measurement of variables: nominal, ordinal, interval, and ratio.

A nominal variable, as the title suggests, is simply a naming variable. More specifically, a variable is said to be nominal if its categories are mutually exclusive and exhaustive. Thus, all variables are at least nominal in level. Common examples of nominal variables in social and behavioral sciences include gender, race, university class (i.e., freshman, sophomore, junior, and senior), and the like.

The next level, the ordinal variable, as the name suggests, adds the requirement of ordered attributes. In other words, a variable is considered ordinal if, in addition to the mutual exclusivity and exhaustive requirements, its attributes can be rank ordered. An important point to note is that the distance between ranked attributes of an ordinal variable can be quite disparate. For example, the commonly used ordinal measure of self-rated health, which typically includes the ordered attributes “poor,” “fair,” “good,” “very good,” and “excellent,” consists of attributes that probably are spaced in unequal intervals (i.e., it is likely that the distance between the “excellent” and “very good” categories is quite different from that between “poor” and “fair” even though the differences in ranks are the same). Ordinal measures are frequently found in surveys where attitudes or behaviors are assessed using measures of levels of agreement (e.g., Likert scaling) or frequencies (e.g., categories ranging from “never” to “always”).

The third level, the interval variable, has variable attributes that also can be ordered. However, unlike the ordinal measure, the spacing between attributes in an interval-level variable is equal. Thus, distances between variable attributes can be compared directly. For example, temperature in degrees Fahrenheit is a common textbook example of an interval-level variable. With this measure, one not only can indicate that 80 degrees is warmer than 40 degrees, but also it is valid to say that the distance from 40 to 60 degrees is the same as the distance from 60 to 80 degrees. However, because interval measures do not include a “true” zero point, one cannot make statements such as “80 degrees is twice as warm as 40 degrees.” Interval-level measures are the least common of the four measurement levels found in the social and behavioral sciences and typically consist of created scales (e.g., IQ).

The final measurement level, the ratio level, differs from the interval level only in that variables in this level have a “true” zero point among their attributes. Thus, the attributes of such variables can be expressed as ratios. For example, income measured in U.S. dollars can be validly expressed in the form of a ratio. A subject with an annual income of $40,000 can be said to have twice the annual income of another subject who reported only $20,000. Ratio-level variables provide the most detailed information of any variable type allowing for the examination of the most complex research questions; however, obtaining this level of detailed measurement is often impractical because of difficulties in measuring social and behavioral phenomena at this level of precision as well as the expense of such detailed measurement.

Variables in Analyses

Although the attributes and measurement levels of variables describe the nature of such measures, variables are frequently referred to by terms related to their analytic use. These terms can be divided into two separate foci: (1) the conceptual use of variables in research design and theory testing and (2) the practical use of variables in statistical analyses.

Conceptual Use

In research design, it is necessary to translate theoretical assertions regarding "causes" and "effects" into hypotheses for scientific testing. In doing so, an additional vocabulary is used to describe variables in this conceptual logic. Depending on the particular research design, these terms differ subtly.

In a basic experimental design, the "cause" is usually termed the treatment variable and refers to the measure that differs systematically across randomly assigned groups. The "effect" is translated into an outcome variable, which refers to the result of theoretical interest (e.g., in a clinical trial, the presence or absence of a medical condition). Additionally, a placebo variable, which measures the effect of the experiment itself on subjects, might be used as well as block variables that consider elements of the research design that cannot be randomized.

On the other hand, the language that is used to describe variables in survey research is somewhat different and reflects the analytic techniques typically used to analyze data from such research designs. In survey designs, the "cause" is usually called the independent variable, and almost always, there will be multiple independent variables in any set of analyses. The "effect" is usually termed the dependent variable, although it is sometimes referred to as the outcome variable just as in experimental research. Because survey designs rely on random sampling rather than on random subject assignment to the "treatment" or "control" categories of a treatment variable, another type of variable, the control variable, is usually incorporated into analyses. The use of control variables is an (imperfect) approximation of the control group assignment procedure where variables that are believed to affect the outcome of interest are included in statistical analyses to eliminate possible spurious variable associations. Typically, control variables in behavioral and social research consist of some combination of demographic measures (e.g., gender, race, age, etc.) and/or measures shown in previous research to affect the outcome under study.

Practical Use

Once theoretical concepts have been translated into observed measurements, analyses are completed to test theoretical assertions. In applying the appropriate analyses, variables are discussed with yet another set of terms.

For the purposes of statistical analysis, variables can be divided into two groupings: categorical variables and numerical variables. Categorical variables include any variables measured at either the nominal or the ordinal level. Numerical variables include any variables measured at the interval or ratio level. Analyses that use variables only at the nominal or ordinal levels of measurement are limited to simple descriptive statistics because of the nonnumeric nature of such measures, whereas the full set of arithmetic operations is possible with the variables measured at the interval or ratio levels. Frequently, however, analyses include both categorical and numerical variables, for example, by examining either categorical group differences across a numerical variable (e.g., means testing) or through statistical modeling procedures. When statistical modeling is employed, a special type of categorical variable, the indicator variable (or "dummy" variable), is often used. Simply put, an indicator variable includes only two categories that indicate having or not having a particular value (e.g., being female or not female). Any categorical variable, regardless of the number of categories defined, can be decomposed into sets of indicator variables.

A final note of caution is warranted. It has become common practice to use the terms *discrete variable* and *continuous variable* synonymously with categorical and numerical variables, respectively. Technically, however, this is incorrect. Discrete variables are measures that can take on only a finite set of values, such as counts. Continuous variables, on the other hand, might consist of any value within a specified range. Thus, discrete and continuous variables are merely subtypes of the numerical variable. However, one can consider the indicator variable, although technically a categorical variable, also to be a discrete variable by noting that its two categories represent a finite set of numerical values (i.e., all or none). Indeed, it is this property of the indicator variable that allows for its use in statistical modeling.

- levels of measurement
- statistical models
- variables (research)
- experimental designs
- research design
- measurement
- control variable

J. Scott Brown
http://dx.doi.org/10.4135/9781412961288.n490
See also

- Categorical Variable
- Control Variables
- Dependent Variable
- Dichotomous Variable
- Independent Variable
- Interval Scale
- Levels of Measurement
- Nominal Scale
- Ordinal Scale
- Predictor Variable
- Ratio Scale

Further Readings
Babbie, E. (2007). The practice of social research (11th ed). Belmont, CA: Thomson Wadsworth.
Blalock, H. M., Jr. (1972). Social statistics. New York: McGraw-Hill.
Hatcher, L. (2003). Step-by-step basic statistics using SAS. Cary, NC: SAS Institute.
Krazanowski, W. (2007). Statistical principles and techniques in scientific and social research. New York: Oxford University Press.
Miller, D. C. (1991). Handbook of research design and social measurement (5th ed). Newbury Park, CA: Sage.