Encyclopedia of Research Design

Independent Variable

Contributors: Shihe Fan
Edited by: Neil J. Salkind
Book Title: Encyclopedia of Research Design
Chapter Title: "Independent 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.n184
Print pages: 592-593

Independent variable is complementary to dependent variable. These two concepts are used primarily in their mathematical sense, meaning that the value of a dependent variable changes in response to that of an independent variable. In research design, independent variables are those that a researcher can manipulate, whereas dependent variables are the responses to the effects of independent variables. By purposefully manipulating the value of an independent variable, one hopes to cause a response in the dependent variable.

As such, independent variables might carry different names in various research fields, depending on how the relationship between the independent and the dependent variable is defined. They might be called explanatory variables, controlled variables, input variables, predictor variables, factors, treatments, conditions, or other names. For instance, in regression experiments, they often are called regressors in relation to the regressand, the dependent, or the response variable.

The concept of independent variable in statistics should not be confused with the concept of independent random variable in probability theories. In the latter case, two random variables are said to be independent if and only if their joint probability is the product of their marginal probabilities for every pair of real numbers taken by the two random variables. In other words, if two random variables are truly independent, the events of one random variable have no relationship with the events of the other random variable. For instance, if a fair coin is flipped twice, a head occurring in the first flip has no association with whether the second flip is a head or a tail because the two events are independent.

Mathematically, the relationship between independent and dependent variables might be understood in this way:

$$y = f(x), \qquad (1)$$

where *x* is the independent variable (i.e., any argument to a function) and *y* is the dependent variable (i.e., the value that the function is evaluated to). Given an input of *x*, there is a corresponding output of *y,x* changes independently, whereas *y* responds to any change in *x*.

Equation 1 is a deterministic model. For each input in *x*, there is one and only one response in *y*. A familiar graphic example is a straight line if there is only one independent variable of order 1 in the previous model. In statistics, however, this model is grossly inadequate. For each value of *x*, there is often a population of *y*, which follows a probability distribution. To reflect more accurately this reality, the preceding equation is revised accordingly:

$$E(y) = f(x), \qquad (2)$$

where *E*(*y*) is the expectation of *y*, or equivalently,

$$y = f(x) + \varepsilon, \qquad (3)$$

where ε is a random variable, which follows a specific probability distribution with a zero mean. This is a probabilistic model. It is composed of a deterministic part [*f*/(x)] and a random part (ε). The random part is the one that accounts for the variation in *y*.

In experiments, independent variables are the design variables that are predetermined by researchers before an experiment is started. They are carefully controlled in controlled experiments or selected in observational studies (i.e., they are manipulated by the researcher according to the purpose of a study). The dependent variable is the effect to be observed and is the primary interest of the study. The value of the dependent variable varies subjecting to the variation in the independent variables and cannot be manipulated to establish an artificial relationship between the independent and dependent variables. Manipulation of the dependent variable invalidates the entire study.

Because they are controlled or preselected and are usually not the primary interest of a study, the value of independent variables is almost universally not analyzed. Instead, they are simply taken as prescribed. (This, however, does not preclude the numeric description of independent variables as can be routinely seen in scientific literature. In fact, they are often described in detail so that a published study can be evaluated properly or repeated by others.) In contrast, the value of the dependent variable is unknown before a study. The observed value of the dependent variable usually requires careful analyses and proper explanation after a study is done.

A caution note must be sounded that even though the value of independent variables can be manipulated, one should not change it in the middle of a study. Doing so drastically modifies the independent variables before and after the change, causing a loss of the internal validity of the study. Even if the value of the dependent variable does not change drastically in response to a manipulation such, the result remains invalid. Careful selection and control of independent variables before and during a study is fundamental to both the internal and the external validity of that study.

To illustrate what constitutes a dependent variable and what is an independent variable, let us assume an agricultural experiment on the productivity of two wheat varieties that are grown under identical or similar field conditions. Productivity is measured by tons of wheat grains produced per season per hectare. In this experiment, variety would be the independent variable and productivity the dependent variable. The qualifier, “identical or similar field conditions,” implies other extraneous (or nuisance) factors (i.e., covariates) that must be controlled, or taken account of, in order for the results to be valid. These other factors might be the soil fertility, the fertilizer type and amount, irrigation regime, and so on. Failure to control or account for these factors could invalidate the experiment. This is an example of controlled experiments. Similar examples of controlled experiments might be the temperature effect on the hardness of a type of steel and the speed effect on the crash result of automobiles in safety tests.

Consider also an epidemiological study on the relationship between physical inactivity and obesity in young children: The parameter(s) that measures physical inactivity, such as the hours spent on watching television and playing video games, and the means of transportation to and from daycares/ schools is the independent variable. These are chosen by the researcher based on his or her preliminary research or on other reports in literature on the same subject prior to the study. The parameter(s) that measure obesity, such as the body mass index, is (are) the dependent variable. To control for confounding, the researcher needs to consider, other than the main independent variables, any covariate that might influence the dependent variable. An example might be the social economical status of the parents and the diet of the families.

Independent variables are predetermined factors that one controls and/or manipulates in a designed experiment or an observational study. They are design variables that are chosen to incite a response of a dependent variable. Independent variables are not the primary interest of the experiment; the dependent variable is.

Shihe Fan
http://dx.doi.org/10.4135/9781412961288.n184

See also

- Bivariate Regression
- Covariate
- Dependent Variable

Further Readings

Hockling, R. R. (2003). Methods and applications of linear models: Regression and the analysis of variance. Hoboken, NJ: Wiley.
Kuehl, R. O. (1994). Statistical principles of research design and analysis. Belmont, CA: Wadsworth.
Montgomery, D. C. (2001). Design and analysis of experiments (
5th ed.
). Toronto, Ontario, Canada: Wiley.
Ramsey, F. L., & Schafer, D. W. (2002). The statistical sleuth: A course in methods of data analysis (
2nd ed.
). Pacific Grove, CA: Duxbury.
Wacherly, D. D., Mendenhall, W., III, & Scheaffer, R. L. (2002). Mathematical statistics with applications (
6th ed.
). Pacific Grove, CA: Duxbury.
Zolman, J. F. (1993). Biostatistics: Experimental design and statistical inference. New York: Oxford University Press.