Encyclopedia of Research Design

Control Variables

Contributors: Jason D. Pole & Susan J. Bondy
Edited by: Neil J. Salkind
Book Title: Encyclopedia of Research Design
Chapter Title: "Control Variables"
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.n77
Print pages: 253-254

In experimental and observational design and data analysis, the term *control variable* refers to variables that are not of primary interest (i.e., neither the exposure nor the outcome of interest) and thus constitute an extraneous or third factor whose influence is to be controlled or eliminated. The term refers to the investigator's desire to estimate an effect (such as a measure of association) of interest that is independent of the influence of the extraneous variable and free from bias arising from differences between exposure groups in that third variable.

Extraneous variables of this class are usually those variables described as potential confounders in some disciplines. Controlling for a potential confounder, which is not an effect modifier or mediator, is intended to isolate the effect of the exposure of interest on the outcome of interest while reducing or eliminating potential bias presented by differences in the outcomes observed between exposed and unexposed individuals that are attributable to the potential confounder. Control is achieved when the potential confounder cannot vary between the exposure groups, and thus the observed relationship between the exposure and outcome of interest is independent of the potential confounder.

As an example, if an investigator is interested in studying the rate of a chemical reaction (the outcome) and how it differs with different reagents (the exposures), the investigator may choose to keep the temperature of each reaction constant among the different reagents being studied so that temperature differences could not affect the outcomes.

Potential control variables that are mediators in another association of interest, as well as potential control variables that are involved in a statistical reaction with other variables in the study, are special cases which must be considered separately. This entry discusses the use of control variables during the design and analysis stages of a study.

Design Stage

There are several options for the use of control variables at the design stage. In the example about rates of reaction mentioned earlier, the intention was to draw conclusions, at the end of the series of experiments, regarding the relationship between the reaction rates and the various reagents. If the investigator did not keep the temperature constant among the series of experiments, difference in the rate of reaction found at the conclusion of the study may have had nothing to do with different reagents, but be solely due to differences in temperature or some combination of reagent and temperature. Restricting or specifying a narrow range of values for one or more potential confounders is frequently done in the design stage of the study, taking into consideration several factors, including ease of implementation, convenience, simplified analysis, and expense. A limitation on restriction may be an inability to infer the relationship between the restricted potential confounder and the outcome and exposure. In addition, residual bias may occur, owing to incomplete control (referred to as *residual confounding*).

Matching is a concept related to restriction. Matching is the process of making the study group and control group similar with regard to potential confounds. Several different methods can be employed, including *frequency matching, category matching, individual matching, and caliper matching*. As with restriction, the limitations of matching include the inability to draw inferences about the control variable(s). Feasibility can be an issue, given that a large pool of subjects may be required to find matches. In addition, the potential for residual confounding exists.

Both matching and restriction can be applied in the same study design for different control variables.

The Analysis Stage

There are several options for the use of control variables at the analysis stage. Separate analysis can be undertaken for each level of a potential confounder. Within each unique value (or homogeneous stratum) of the potential confounder, the relationship of interest may be observed that is not influenced by differences between exposed and unexposed individuals attributable to the potential confounder. This technique is another example of restriction.

Estimates of the relationship of interest independent of the potential confounder can also be achieved by the use of a matched or stratified approach in the analysis. The estimate of interest is calculated at all levels (or several theoretically homogeneous or equivalent strata) of the potential confounder, and a weighted, average effect across strata is estimated. Techniques of this kind include the *Mantel–Haenszel stratified analysis*, as well as stratified (also called matched or conditional) regression analyses. These approaches typically assume that the stratum-specific effects are not different (i.e., no effect modification or statistical interaction is present). Limitations of this method are related to the various ways strata can be formed for the various potential confounders, and one may end up with small sample sizes in many strata, and therefore the analysis may not produce a reliable result.

The most common analytic methods for using control variables is analysis of covariance and multiple generalized linear regression modeling. Regression techniques estimate the relationship of interest conditional on a fixed value of the potential confounder, which is analogous to holding the value of the potential confounder constant at the level of third variable. By default, model parameters (intercept and beta coefficients) are interpreted as though potential confounders were held constant at their zero values. Multivariable regression is relatively efficient at handling small numbers and easily combines variables measured on different scales.

Where the potential control variable in question is involved as part of a statistical interaction with an exposure variable of interest, holding the control variable constant at a single level through restriction (in either the design or analysis) will allow estimation of the effect of the exposure of interest and the outcome that is independent of the third variable, but the effect measured applies only to (or is conditional on) the selected level of the potential confounder. This would also be the stratum-specific or conditional effect. For example, restriction of an experiment to one gender would give the investigator a gender-specific estimate of effect.

If the third variable in question is part of a true interaction, the other forms of control, which permit multiple levels of the third variable to remain in the study (e.g., through matching, statistical stratification, or multiple regression analysis), should be considered critically before being applied. Each of these approaches ignores the interaction and may serve to mask its presence.

- control variable
- regression
- temperature
- residuals

- regression analysis
- estimates
- control groups

Jason D. Pole, and Susan J. Bondy
http://dx.doi.org/10.4135/9781412961288.n77

See also

- Bias
- Confounding
- Interaction
- Matching

Further Readings

Cook, T. D., & Campbell, D. T. (1979). Quasi-experimentation: Design and analysis issues for field settings. Boston: Houghton Mifflin.