

Control

- Data analysis consists of estimating slope coefficients with and without particular variables “controlled”
- Random assignment will show only random differences for any prior variable
- If prior variables can be measured, they can be controlled for by statistical adjustment
 - Residualization in multiple regression
 - Replication with control categories in tabular analysis
 - Standardization or weighting in demography
- Bivariate relationship—the relationship between X_i and X_j with no variables controlled
- Partial relationship—the relationship between X_i and X_j with one or more variables controlled

Rules of Causal Order

RULE #7

When estimating the effect of X_i on X_j , control all prior and intervening variables; i.e., control all variables not consequent to X_j

RULE #7a

In multiple regression, regress each variable on all its predecessors

Lazarfeldian Elaboration

- Complicated matters are compressed into two numbers
 - The bivariate relationship
 - The partial relationship

RULE #8

A test variable, X_t , explains $X_i X_j$ when the $X_i X_j$ bivariate is nonzero and the $X_i X_j$ partial, controlling for X_t , is zero or trivially small. To explain a positive correlation, look for X_t s where $X_t X_i$ and $X_t X_j$ have the same sign. To explain a negative correlation, look for X_t s where $X_t X_i$ and $X_t X_j$ have opposite signs.

- Unless X_t is related to both X_i and X_j it is a poor candidate for explainer.

Hauser & Alwin's Effect Analysis

- A third correlation—the relationship between X_i and X_j controlling for priors only—expands our information considerably
- “effects analysis” or “the decomposition of effects”
- The new coefficient will be the total causal effect of X_i and X_j ; it will be causal because all priors have been controlled, it will be total because both direct effects and indirect effects via interveners have been left in.

RULE #9

In addition to the $X_i X_j$ bivariate and the direct $X_i X_j$ effect, calculate the causal effect by controlling for prior variables only. Then find the spurious components and the indirect portion due to intervening variables.

Sewall Wright's Path Analysis

- Path analysis is limited to specific statistical techniques including multiple regression and structural equation modeling
- The value of a path is found by multiplying the coefficients of each arrow in the path. Regardless of the measurement units of the intervening variables, the result will come out in X_j units per one unit difference in X_i .
- Allows the decomposition of spurious influence into elementary parts so we can see which prior variables are doing what

RULE #10

In a linear system the total causal effect of X_i on X_j is the sum of the values of all the paths from X_i to X_j

Path Analysis (cont'd)

- When looking for an intervening variable to explain a correlation, you need one with strong links to *both* X_i and X_j
- If the paths from X_i to X_j are inconsistent (some positive, some negative), they will add up to the total causal effect, but it is impossible to divide the total into proportions
- Standardized regression coefficient
- Dummy variables

The Cost of Being Wrong

Specification errors are important, as it is possible to obtain results that are statistically impeccable but totally erroneous

- If any control variables are actually consequent to X_j we can do ourselves considerable damage
- Providing X_t s are not consequent to X_j , errors in order have no effect on direct relationships or Lazarfeldian elaboration
- Effect analysis is insensitive to order among priors and order among intervenors
- Absent variables might do anything

Counter-principles in Causal Specification

- There are very few “whopper” effects in social science
- For some important questions (e.g., elaboration) most specification errors have minor or no negative effects
- You can and should use the principles of causal logic to work out for yourself potential sources and consequences of specific error