2004

Grouped Data

Gudmund R. Iversen
Swarthmore College, iversen@swarthmore.edu

The SAGE Encyclopedia of Social Science Research Methods

Grouped Data

Contributors: Gudmund R. Iversen
Edited by: Michael S. Lewis-Beck, Alan Bryman & Tim Futing Liao
Book Title: The SAGE Encyclopedia of Social Science Research Methods
Chapter Title: "Grouped Data"
Pub. Date: 2004
Access Date: March 13, 2020
Publishing Company: SAGE Publications, Inc.
City: Thousand Oaks
Print ISBN: 9780761923633
Online ISBN: 9781412950589
DOI: http://dx.doi.org/10.4135/9781412950589.n383
Print page: 446

Data come from VARIABLES measured on one or more units. In the social sciences, the unit is often an individual. Units used in other sciences could be elements such as a pig, a car, or whatever. It is also possible to have *larger* units, such as a county or a nation. Such units are often made up by aggregating data across individual units. We aggregate the values of a variable across the individual persons and get *grouped data* for the larger unit. Grouped data can consist of data on variables such as average income in a city or number of votes cast in a ward for a candidate.

However, data on aggregates do not always consist of grouped data. If we consider type of government in a country as a variable with values *Democracy, Dictatorship*, and *Other*, then we clearly have data on aggregates of individuals in the countries. But the value for any country is a characteristic of the people who make up the country, and it is not an aggregate of values of any variable across the individuals in that country.

Grouped data exist at various levels of aggregation. There can be grouped data for the individuals who make up a census tract, a country, or a state. The importance of this lies in the fact that aggregation to different levels can produce different results from analyses of the data. Data on different levels may produce different magnitudes of CORRELATIONS depending on whether the data consist of observations aggregated, say, to the level of the county, where we may have data on all 3,000+ counties in the country, or whether the data are aggregated to the level of the state, where we may have data on all 50 states. Thus, analysis results that come from one level of aggregation apply *only* to the level on which the analysis is done. Mostly, they do not apply to units at lower or higher levels of aggregation.

In particular, results obtained from the analysis of aggregate data do in no way necessarily apply to the level of individuals as well. The so-called ECOLOGICAL FALLACY occurs when results obtained from grouped data are thought to apply on the level of the individual as well. In sociology, Robinson (1950) made this very clear in his path-breaking article on this topic. He showed mathematically how an ecological correlation coefficient obtained from grouped data could be very different, both in magnitude and in sign, from the correlation of the same two variables using data on individuals. *Simpson's paradox* is another name used for this phenomenon.

This situation becomes worse when there are group data available, but the individual data have not been observed. In voting, we know the number of votes cast for a candidate as well as other characteristics of precincts, but we do not know how the single individuals voted. Attempts have been made to construct methods to recover underlying individual-level data from group data, but in principle, such recovery will remain impossible without additional information.

In social science data analysis, another common form of grouped data is cross-classified data (i.e., observations cross-classified by the categories of the variables in an analysis). Such cross-classifications are known as CONTINGENCY TABLES and are often analyzed by LOG-LINEAR MODEL.

Gudmund R. Iversen
http://dx.doi.org/10.4135/9781412950589.n383
See also

- DATA

References

Borgatta, E. F., & Jackson, D. J. (Eds.). (1980). Aggregate data: Analysis and interpretation. Beverly Hills, CA: Sage.
Iversen, G. R.Recovering individual data in the presence of group and individual effects. American Journal of Sociology79420–434(1973).
Robinson, W. S.Ecological correlations and the behavior of individuals. American Sociological Review15351–357(1950).