2.1F: Analyzing Data and Drawing Conclusions

Data analysis in sociological research aims to identify meaningful sociological patterns.

Learning Objectives

• Compare and contrast the analysis of quantitative vs. qualitative data

Key Points

• Analysis of data is a process of inspecting, cleaning, transforming, and modeling data with the goal of highlighting useful information, suggesting conclusions, and supporting decision making. Data analysis is a process, within which several phases can be distinguished.

• One way in which analysis can vary is by the nature of the data. Quantitative data is often analyzed using regressions. Regression analyses measure relationships between dependent and independent variables, taking the existence of unknown parameters into account.

• Qualitative data can be coded—that is, key concepts and variables are assigned a shorthand, and the data gathered are broken down into those concepts or variables. Coding allows sociologists to perform a more rigorous scientific analysis of the data.

• Sociological data analysis is designed to produce patterns. It is important to remember, however, that correlation does not imply causation; in other words, just because variables change at a proportional rate, it does not follow that one variable influences the other.

• Without a valid design, valid scientific conclusions cannot be drawn. Internal validity concerns the degree to which conclusions about causality can be made. External validity concerns the extent to which the results of a study are generalizable.
Key Terms

- **correlation**: A reciprocal, parallel or complementary relationship between two or more comparable objects.
- **causation**: The act of causing; also the act or agency by which an effect is produced.
- **Regression analysis**: In statistics, regression analysis includes many techniques for modeling and analyzing several variables, when the focus is on the relationship between a dependent variable and one or more independent variables. More specifically, regression analysis helps one understand how the typical value of the dependent variable changes when any one of the independent variables is varied, while the other independent variables are held fixed.

The Process of Data Analysis

Analysis of data is a process of inspecting, cleaning, transforming, and modeling data with the goal of highlighting useful information, suggesting conclusions, and supporting decision making. In statistical applications, some people divide data analysis into descriptive statistics, exploratory data analysis (EDA), and confirmatory data analysis (CDA). EDA focuses on discovering new features in the data and CDA focuses on confirming or falsifying existing hypotheses. Predictive analytics focuses on the application of statistical or structural models for predictive forecasting or classification. Text analytics applies statistical, linguistic, and structural techniques to extract and classify information from textual sources, a species of unstructured data.

Data analysis is a process, within which several phases can be distinguished. The initial data analysis phase is guided by examining, among other things, the quality of the data (for example, the presence of missing or extreme observations), the quality of measurements, and if the implementation of the study was in line with the research design. In the main analysis phase, either an exploratory or confirmatory approach can be adopted. Usually the approach is decided before data is collected. In an exploratory analysis, no clear hypothesis is stated before analyzing the data, and the data is searched for models that describe the data well. In a confirmatory analysis, clear hypotheses about the data are tested.

Regression Analysis

The type of data analysis employed can vary. One way in which analysis often varies is by the quantitative or qualitative nature of the data.

Quantitative data can be analyzed in a variety of ways, regression analysis being among the most popular. Regression analyses measure relationships between dependent and independent variables, taking the existence of unknown parameters into account. More specifically, regression analysis helps one understand how the typical value of the dependent variable changes when any one of the independent variables is varied, while the other independent variables are held fixed.
Linear Regression: This graph illustrates random data points and their linear regression.

A large body of techniques for carrying out regression analysis has been developed. In practice, the performance of regression analysis methods depends on the form of the data generating process and how it relates to the regression approach being used. Since the true form of the data-generating process is generally not known, regression analysis often depends to some extent on making assumptions about this process. These assumptions are sometimes testable if a large amount of data is available. Regression models for prediction are often useful even when the assumptions are moderately violated, although they may not perform optimally. However, in many applications, especially with small effects or questions of causality based on observational data, regression methods give misleading results.

Coding

Qualitative data can involve coding—that is, key concepts and variables are assigned a shorthand, and the data gathered is broken down into those concepts or variables. Coding allows sociologists to perform a more rigorous scientific analysis of the data. Coding is the process of categorizing qualitative data so that the data becomes quantifiable and thus measurable. Of course, before researchers can code raw data such as taped interviews, they need to have a clear research question. How data is coded depends entirely on what the researcher hopes to discover in the data; the same qualitative data can be coded in many different ways, calling attention to different aspects of the data.

Coding Qualitative Data: Qualitative data can be coded, or sorted into categories. Coded data is quantifiable. In this bar chart, help requests have been coded and categorized so we can see which types of help requests are most
Sociological Data Analysis

Sociological data analysis is designed to produce patterns. It is important to remember, however, that correlation does not imply causation; in other words, just because variables change at a proportional rate, it does not follow that one variable influences the other.

**Correlation, Causation, and Spurious Relationships**: This mock newscast gives three competing interpretations of the same survey findings and demonstrates the dangers of assuming that correlation implies causation.

Conclusions

In terms of the kinds of conclusions that can be drawn, a study and its results can be assessed in multiple ways. Without a valid design, valid scientific conclusions cannot be drawn. Internal validity is an inductive estimate of the degree to which conclusions about causal relationships can be made (e.g., cause and effect), based on the measures used, the research setting, and the whole research design. External validity concerns the extent to which the (internally valid) results of a study can be held to be true for other cases, such as to different people, places, or times. In other words, it is about whether findings can be validly generalized. Learning about and applying statistics (as well as knowing their limitations) can help you better understand sociological research and studies. Knowledge of statistics helps you makes sense of the numbers in terms of relationships, and it allows you to ask relevant questions about sociological phenomena.