Lecture Notes

# Chapter 1: The What and the Why of Statistics

## Learning Objectives

1. Describe the five stages of the research process.
2. Define independent and dependent variables.
3. Distinguish between the three levels of measurement.
4. Apply descriptive and inferential statistical procedures.

## Chapter Outline

1. Introduction
   1. Statistics is not just a course but a useful tool.
   2. Reasons why learning statistics is important:
      1. By learning statistics, one will become a sharper consumer of statistical material.
      2. A person may be expected to read and interpret statistical information related to their occupation or work.
         1. Even if conducting research is not a part of a person’s work, he or she may still be expected to understand and learn from other people’s research or to be able to write reports based on statistical analyses.
   3. **Statistics** refers to a set of procedures used by social scientists to organize, summarize, and communicate numerical information.
   4. Information represented by numbers is called **data;** researchers use statistical procedures to analyze data to answer research questions and test theories.
2. The Research Process
   1. The **research process** is a set of activities in which social scientists engage so that they can answer questions, examine ideas, or test theories.
   2. The research process consists of five stages:
      1. Asking the research question,
      2. Formulating the hypotheses,
      3. Collecting data,
      4. Analyzing data, and
      5. Evaluating the hypotheses.
   3. Each stage affects the theory and is affected by it as well.
   4. Statistics is most closely tied to the data analysis stage of the research process.
   5. Statistical analysis of the data helps researchers test the validity and accuracy of their hypotheses.
3. Asking Research Questions
   1. The starting point for most research is asking a research question that can be answered by conducting **empirical research,** research based on information that can be verified from direct experience.
   2. We study these questions by defining good and better in terms that can be verified empirically and then answer these questions by conducting empirical research.
   3. To come up with a research question:
      1. Pick a question that interests you.
      2. Talk to other people.
      3. Write down your own observations and ideas OR
      4. Learn what other social scientists have written about.
4. The Role of Theory
   1. A research question is expressed in terms of a relationship between two or more attributes of individuals or groups.
   2. A **theory** is a set of assumptions and propositions used by social scientists to explain, predict, and understand the phenomena they study.
   3. It attempts to establish a link between what is observed (the data) and conceptual understanding of why certain phenomena are related to each other in a particular way.
5. Formulating the Hypothesis
   1. Theories suggest specific concrete predictions or **hypothesis** (a statement predicting the relationship between two or more observable attributes) about the way that observable attributes of people or groups are interrelated in real life.
   2. Hypotheses are tentative because they can be verified only after they have been tested empirically.
   3. We can generate hypotheses from theories, directly from observations, from intuition or from professional or scholarly literature.
   4. A **variable** is a property of people or objects that takes on two or more values.
   5. Social scientists must also select a **unit of analysis**, the object of their research.
   6. Independent and Dependent Variables: Causality
      1. Hypotheses are usually stated in terms of a relationship between an independent and a dependent variable.
         1. In the language of research, the variable the researcher wants to explain (“the effect”) is called the **dependent variable.**
         2. The variable that is expected to “cause” or account for the dependent variable is called the **independent variable**.
      2. Cause-and-effect relationships between variables are not easy to infer in the social sciences.
6. To establish that two variables are causally related, analysis must meet three conditions:
   * + - 1. The cause has to precede the effect in time.
         2. There has to be an empirical relationship between the cause and the effect.
         3. This relationship cannot be explained by other factors.
   1. Independent and Dependent Variables: Guidelines
      1. It is difficult to infer cause-and-effect relationships in the social sciences.
      2. Using the terms independent variable and dependent variable is still appropriate even when this relationship is not articulated in terms of direct cause and effect.
      3. A few guidelines that may help identify the independent and dependent variables:
         1. The dependent variable is always the property that you are trying to explain; it is always the object of the research.
         2. The independent variable usually occurs earlier in time than the dependent variable.
         3. The independent variable is often seen as influencing, directly or indirectly, the dependent variable.
      4. The purpose of the research is to determine which the independent variable is and which the dependent variable is.
      5. A variable defined as independent in one research investigation may be a dependent variable in another.
7. Collecting Data
   1. After deciding the research question, the hypothesis, and the variables to be included in the study, we proceed to measuring our variables and collecting the data.
   2. We must decide:
      1. How to measure the variables of interest to us.
      2. How to select the cases for our research.
      3. What kind of data collection techniques we will be using.
   3. A wide variety of data collection techniques are available to us, from direct observations to survey research, experiments, or secondary sources.
   4. We can construct numerous measuring instruments ranging in complexity from a single question included in a questionnaire to composite measure constructed through the combination of two or more questionnaire items.
   5. The choice of a particular data collection method or instrument to measure our variables depends on the study objective.
   6. Levels of Measurement
      1. The type of statistical operation we employ depends on how our variables are measured.
      2. The correspondence between the properties we measure and the numbers representing these properties determines the type of statistical operations we can use.
      3. The degree of correspondence also leads to distinct levels of measurement.
      4. Nominal Level of Measurement
         1. Numbers or other symbols are assigned a set of categories for the purpose of naming, labeling, or classifying the observations.
         2. We could use any of a variety of symbols to represent the different categories of a nominal variable.
         3. When numbers are used to represent the different categories, we do not imply anything about the magnitude or quantitative difference between the categories.
         4. Nominal categories cannot be rank-ordered.
         5. Because the different categories vary in the quality inherent in each but not in quantity, nominal variables are often called qualitative.
         6. Nominal variables should include categories that are both exhaustive and mutually exclusive.
         7. Mutual exclusiveness means that there is only one category suitable for each observation.
            1. This is not recommended if it leads to the exclusion of categories that have theoretical significance or a substantial number of observations.
      5. Ordinal Level of Measurement
8. **Ordinal** level of measurement: Numbers are assigned to rank-ordered categories ranging from low to high.
9. Many attitudes that we measure in the social sciences are ordinal-level variables.
10. Ordinal variables should include categories that are mutually exhaustive and exclusive.
    * 1. Interval-Ratio Level of Measurement
11. **Interval ratio** level of measurement: The categories (or values) of a variable can be rank-ordered and the measurements for all the cases are expressed in the same units and equally spaced.
12. With these variables, we can compare values not only in terms of which is larger or smaller but also in terms of how much larger or smaller one is compared with another.
13. In some discussions of levels of measurement, one can see a distinction made between interval-ratio variables that have a natural zero point (where zero means the absence of the property) and those variables that have zero as an arbitrary point.
    * 1. Cumulative Property of Levels of Measurement
14. Variables that can be measured at the interval-ratio level of measurement can also be measured at the ordinal and nominal levels.
15. As a rule, properties that can be measured at a higher level (interval ratio is the highest) can also be measured at lower levels but not vice versa.
    * 1. Levels of Measurement of Dichotomous Variables
16. A variable that has only two values is called a **dichotomous variable.**
17. They may seem to be measured at the nominal level.
18. No category is naturally higher or lower than the other, so they can’t be ordered.
19. Because there are only two possible values for a dichotomy, we can measure it at the ordinal or the interval-ratio level.
    * + - 1. With only two classes, there is no way to get them out of order; therefore, gender could be considered at the ordinal level.
          2. In measuring interval-ratio data, the size of the interval between the categories is meaningful.
20. Researchers often dichotomize some of their variables, turning a multi-category nominal variable into a dichotomy.
    * + - 1. They should make sure that the two categories capture a distinction that is important to the research question.
    1. Discrete and Continuous Variables
       1. Discrete variables have a minimum-sized unit of measurement that cannot be subdivided.
       2. Continuous variables do not have a minimum-sized unit of measurement; their range of values can be subdivided into increasingly smaller fractional values.
       3. Whether variables are continuous or discrete affects subsequent research operations, particularly measurement procedures, data analysis, and methods of inference and generalization.
       4. In practice, some discrete variables can be treated as if they were continuous and vice versa.
21. Analyzing the Data and Evaluating the Hypothesis
    1. After data collection, researchers analyze their data and evaluate the hypotheses of the study.
       1. The data consist of codes and numbers used to represent the observations.
       2. The typical research project includes many variables; therefore, the amount of data the researcher confronts is considerably large.
       3. We must find a systematic way to organize these data, analyze them, and use some set of procedures to decide what they mean.
          1. These last steps make up the statistical analysis stage.
          2. At this point in the research cycle, statistical procedures will help us evaluate our research hypothesis and assess the theory from which the hypothesis was derived.
    2. Descriptive and Inferential Statistics
       1. A **population** is the total set of individuals, objects, groups, or events in which the researcher is interested.
          1. Because of limited time and resources, it is impossible to study the entire population.
       2. A subset of cases selected from a population is called a **sample.**
22. Researchers usually collect their data from a sample and then generalize their observations to the population.
    * 1. The process of identifying and selecting this subset is referred to as **sampling.**
23. The ultimate goal of sampling is to have a subset that closely resembles the characteristics of the population.
    * 1. **Descriptive statistics** includes procedures that help us organize and describe data collected from either a sample or a population.
24. Occasionally data are collected on an entire population, as in a census.
    * 1. **Inferential statistics** makes predictions or inferences about a population based on observations and analyses of a sample.
25. Because the data are based on a sample rather than on the entire population, the average of the sample does not equal the average of the population as a whole.
    1. Evaluating the Hypotheses
       1. Next stage of the research process is the assessment and evaluation of our hypotheses and theories in light of the analyzed data.
       2. New questions might be raised about unexpected trends in the data and about other variables that may have to be considered in addition to our original variables.
       3. These findings provide evidence to help us decide how our data relate to the theoretical framework that guided our research.
       4. We may decide to revise our theory and hypothesis to take into account of these later findings.
       5. Statistics helps us evaluate our hypotheses and theories, discover unanticipated patterns and trends, and provide the impetus for shaping and reformulating our theories.
26. Examining a Diverse Society
    1. Less partial and distorted explanations of social relations tend to result when researchers, research participants, and the research process itself reflect the increasing diversity of American society.
    2. A consciousness of social differences shapes the research questions we ask, how we observe and interpret our findings, and the conclusions we draw.
    3. Statistical procedures and quantitative methodologies can be used to describe our diverse society.
27. Learning Statistics
    1. Students learning statistics should understand the vocabulary of statistics and be able to translate the symbols and codes into familiar terms.
    2. Another strategy for increasing one’s statistical knowledge is to frame new learning in a context that is relevant and interesting.
    3. To develop confidence in your statistical ability, collaborate with your peers as you learn this course material.