Chapter 1: Introduction to Statistics

Chapter Outline

1.1 Statistics, Science, and Observation
- Introduction
- Definitions of Statistics
- Populations and Samples
- Variables and Data
- Parameters and Statistics
- Descriptive and Inferential Statistical Methods
- Statistics in the Context of Research

1.2 Variables and Measurement
- Constructs and Operational Definitions
- Discrete and Continuous Variables
- Scales of Measurement

1.3 Three Data Structures, Research Methods, and Statistics
- Data Structure 1. One Group with One or More Separate Variables Measured for Each Individual: Descriptive Research
- Relationships between Variables
- Data Structure 2. One Group with Two Variables Measured for Each Individual: The Correlational Method
- Data Structure 3. Comparing Two (or More) Groups of Scores: Experimental and Nonexperimental Methods
- Experimental and Nonexperimental Methods
- The Experimental Method
- Nonexperimental Methods: Nonequivalent Groups and Pre-Post Studies

1.4 Statistical Notation
- Scores
- Summation Notation

Learning Objectives and Chapter Summary

1. Students should be familiar with the terminology and special notation of statistical analysis.

Statistical Terms	Measurement Terms	Research Terms
population	operational definition	correlational method
sample	nominal	experimental method
parameter	ordinal	independent variable
statistic	interval	dependent variable
descriptive statistics	ratio	nonexperimental method
inferential statistics	discrete variable	quasi-independent variable
sampling error	continuous variable	
	real limits	

Figure 1.1 is useful for introducing the concepts of population and sample, and the related concepts of parameter and statistic. The same figure also helps differentiate descriptive statistics that focus on the sample data and inferential statistics that generalize from samples to populations.

2. Students should learn how statistical techniques fit into the general process of science.

Although the concept of sampling error is not critical at this time in the course, it is a useful way to introduce and justify the need for inferential statistics. Figure 1.2 is a simple demonstration of the concept that sample statistics are representative of but not identical to the corresponding population parameters, and that two different samples will tend to have different statistics. The idea that differences can occur just by chance is an important concept. After the concept of sampling error is established, Figure 1.3 shows the overall research process and identifies where descriptive and inferential statistics are used.

Statistical techniques are mostly used near the end of the research process, after the researcher has obtained research results and needs to organize, summarize, and interpret the data. Chapter 1 includes discussion of two aspects of research that precede statistics: (1) the process of measurement, and (2) the idea that measurements take place in the context of a research study. The discussion includes the different scales of measurement and the information they provide, as well as an introduction to continuous and discrete variables. Research studies are described in terms of the kinds of data they produce: correlational studies that produce data suitable for computing correlations (see Figure 1.5), and experimental studies that produce groups of scores to be compared, usually looking for mean differences (see Figure 1.6). Other types of research (nonexperimental) that also involve comparing groups of scores are discussed (see Figure 1.7).

3. Students should learn the notation—particularly summation notation—that will be used throughout the rest of the book.

There are three key concepts important to using summation notation:

1. Summation is a mathematical operation, just like addition or multiplication, and the different mathematical operations must be performed in the correct order (see Order of Mathematical Operations in Section 1.4).
2. In statistics, mathematical operations usually apply to a set of scores that can be presented as a column of numbers.
3. Each operation, except for summation, creates a new column of numbers. Summation calculates the sum for the column.

Other Lecture Suggestions

1. Early in the first class, I acknowledge that

   a. Most students are not there by choice. (No one picked statistics as an elective because it looked like a fun class.)

   b. Many students have some anxiety about the course.

However, I also try to reassure them that the class will probably be easier and more enjoyable (less painful) than they would predict, *provided* they follow a few simple rules:

   a. Keep up. In statistics, each bit of new material builds on the previous material. As long as you have mastered the old material, then the new stuff is just one small step forward. On the other hand, if you do not know the old material, then the new stuff is totally incomprehensible. (For example, try reading Chapter 10 on the first day of class. It will make no sense at all. However, by the time we get to Chapter 10, you will have enough background to understand it.) Keeping up means coming to class, asking questions, and doing homework on a regular basis. If you are getting lost, then get help immediately.

   b. Test yourself. It is very easy to sit in class and watch an instructor work through examples. Also, it is very easy to complete homework assignments if you can look back at example problems in the book. Neither activity means that you really know the material. For each chapter, try one or two of the end-of-chapter problems without looking back at the examples in the book or checking your notes. Can you really do the problems on your own? If not, pay attention to where you get stuck in the problem, so you will know exactly what you still need to learn.

2. Give students a list of variables (for example, items from a survey such as age, gender, education level, income, and occupation), and ask them to identify the scale of measurement most likely to be used and whether the variable is discrete or continuous.

3. Describe a nonexperimental or correlational study and have students identify reasons that you cannot make a cause-and-effect conclusion from the results. For example, a researcher finds that children in the local school who regularly eat a nutritious breakfast have higher grades than students who do not eat a nutritious breakfast. Does this mean that a nutritious breakfast *causes* higher grades? Or a researcher finds that employees who regularly use the company's new fitness center have fewer sick days than employees who do not use the center. Does this mean that using the fitness center *causes* people to be healthier?

For either example, describe how the study could be made into an experiment by:

a. beginning with equivalent groups (random assignment).

b. manipulating the independent variable (note that doing so introduces the ethical question of forcing people to eat a nutritious breakfast).

c. controlling other variables (i.e., the rest of the children's diet).

4. After introducing some basic applications of summation notation, present a simple list of scores (1, 3, 5, 4) and a relatively complex expression containing summation notation, for example, $\Sigma(X - 1)^2$. Ask the students to compute the answer. You are likely to obtain several different responses.

Note that this is not a democratic process—the most popular answer is not necessarily correct. There is only one correct answer because there is only one correct sequence for performing the calculations. Have the class identify the step-by-step sequence of operations specified by the expression. (First, subtract 1 from each of the scores. Second, square the resulting values. Third, sum the squared numbers.) Then apply the steps, one by one, to compute the answer. As a variation, present a list of steps and ask students to write the mathematical expression corresponding to the series of steps.

Answers to Even-Numbered Problems

2. A population is the entire group of individuals of interest. A sample is a relatively small group selected from the population. The research begins with a question about the population. However, the population is usually too large for every individual to participate in the research study. Therefore, the individuals in the sample are the actual participants and the results from the sample are then generalized to the population.

4. A parameter is a characteristic, usually a numerical value, which describes a population. A statistic is a characteristic, usually numerical, that describes a sample. Although samples are generally representative of their populations, they are not perfect. Therefore, there typically is some discrepancy between the statistics from a sample and the corresponding parameters of the population. This naturally occurring discrepancy is called sampling error.

6. Age: ratio scale and continuous. Although people usually report whole-number years, the variable is the amount of time and time is infinitely divisible.

Income: ratio scale and discrete. Income is determined by units of currency. For U.S. dollars, the smallest unit is the penny and there are no intermediate values between 1 cent and 2 cents.

Dependents: ratio scale and discrete. Family size consists of whole-number categories with no intermediate values.

Social Security: nominal scale and discrete. Social Security numbers are essentially names that are coded as 9-digit numbers. There are no intermediate values between two consecutive social security numbers.

8. A correlational study has only one group of individuals and measures two (or more) different variables for each individual. Other research methods evaluating relationships between variables compare two (or more) different groups of scores.

10. This is not an experiment because no independent variable is manipulated. They are comparing two preexisting groups of individuals (those who drink reduced fat milk and those who drink regular or 2% milk).

12. This is not an experiment because there is no manipulation. Instead, the study is comparing two preexisting groups (American and Canadian students).

14. a. The dependent variable is the number of correct answers on the test, which is a measure of knowledge of the material.
    b. Knowledge is a continuous variable. If it is measured with a 10-question test, it may appear to be discrete but it could be measured with a 100-question test, which means that each category can be further divided.

c. ratio scale (zero means none)

16. a. The independent variable is the color of the T-shirt (red versus not-red), and the dependent variable is the size of the tip.

    b. Color of the T-shirt (red versus not-red) is measured on a nominal scale.

18. a. $\Sigma X = 11$

    b. $(\Sigma X)^2 = (11)^2 = 121$

    c. $\Sigma X - 2 = 11 - 2 = 9$

    d. $\Sigma(X - 2) = 3$

20. a. $\Sigma X = -1$

    b. $(\Sigma X)^2 = (-1)^2 = 1$

    c. $\Sigma X^2 = 67$

    c. $\Sigma(X + 3) = 14$

22. a. $(\Sigma X)^2$

    b. ΣX^2

    c. $\Sigma(X - 2)$

    d. $\Sigma(X - 1)^2$