Chapter 1: Introduction to Statistics

**Chapter Outline**

1.1 Statistics, Science, and Observations

 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 Data Structures, Research Methods, and Statistics

 Individual Variables

Relationships between Variables

Statistics for the Correlational Method

Limitations of the Correlational Method

Statistics for Comparing Two (or More) Groups of Scores

Experimental and Nonexperimental Methods

 The Experimental Method

 Terminology in the Experimental Method

Nonexperimental Methods: Nonequivalent Groups and Pre-Post Studies

1.3 Variables and Measurement

Constructs and Operational Definitions

 Discrete and Continuous Variables

 Scales of Measurement

 The Nominal Scale

 The Ordinal Scale

 The Interval and Ratio Scales

1.4 Statistical Notation

 Scores

 Summation Notation

**Learning Objectives and Chapter Summary**

1. Define the terms population, sample, parameter, and statistic, and describe the relationships between them.

The term statistics is used to refer to methods for organizing, summarizing, and interpreting data.

Scientific questions usually concern a population, which is the entire set of individuals one wishes to study. Usually, populations are so large that it is impossible to examine every individual, so most research is conducted with samples. A sample is a group selected from a population, usually for purposes of a research study.

A characteristic that describes a sample is called a statistic, and a characteristic that describes a population is called a parameter. Although sample statistics are usually representative of corresponding population parameters, there is typically some discrepancy between a statistic and a parameter.

1. Define descriptive and inferential statistics and describe how these two general categories of statistics are used in a typical research study.

Statistical methods can be classified into two broad categories: descriptive statistics, which organize and summarize data, and inferential statistics, which use sample data to draw inferences about populations.

1. Describe the concept of sampling error and explain how this concept creates the fundamental problem that inferential statistics must address.

The naturally occurring difference between a statistic and a parameter is called sampling error.

1. Differentiate correlational, experimental, and nonexperimental research and describe the data structures associated with each.
2. Define independent, dependent, and quasi-independent variables and recognize examples of each.
3. Explain why operational definitions are developed for constructs and identify the two components of an operational definition.

The correlational method examines relationships between variables by measuring two different variables for each individual. This method allows researchers to measure and describe relationships, but cannot produce a cause-and-effect explanation for the relationship.

The experimental method examines relationships between variables by manipulating an independent variable to create different treatment conditions and then measuring a dependent variable to obtain a group of scores in each condition. The groups of scores are then compared. A systematic difference between groups provides evidence that changing the independent variable from one condition to another also caused a change in the dependent variable. All other variables are controlled to prevent them from influencing the relationship. The intent of the experimental method is to demonstrate a cause-and-effect relationship between variables. The experimental method examines relationships between variables by manipulating an independent variable to create different treatment conditions and then measuring a dependent variable to obtain a group of scores in each condition. The groups of scores are then compared. A systematic difference between groups provides evidence that changing the independent variable from one condition to another also caused a change in the dependent variable. All other variables are controlled to prevent them from influencing the relationship. The intent of the experimental method is to demonstrate a cause-and-effect relationship between variables.

Nonexperimental studies also examine relationships between variables by comparing groups of scores, but they do not have the rigor of true experiments and cannot produce cause-and-effect explanations. Instead of manipulating a variable to create different groups, a nonexperimental study uses a preexisting participant characteristic (such as male/female) or the passage of time (before/after) to create the groups being compared.

1. Describe discrete and continuous variables and identify examples of each.
2. Differentiate nominal, ordinal, interval, and ratio scales of measurement.

A discrete variable consists of indivisible categories, often whole numbers that vary in countable steps. A continuous variable consists of categories that are infinitely divisible and each score corresponds to an interval on the scale. The boundaries that separate intervals are called real limits and are located exactly halfway between adjacent scores.

A measurement scale consists of a set of categories that are used to classify individuals. A nominal scale consists of categories that differ only in name and are not differentiated in terms of magnitude or direction. In an ordinal scale, the categories are differentiated in terms of direction, forming an ordered series. An interval scale consists of an ordered series of categories that are all equal-sized intervals. With an interval scale, it is possible to differentiate direction and magnitude (or distance) between categories. Finally, a ratio scale is an interval scale for which the zero point indicates none of the variable being measured. With a ratio scale, ratios of measurements reflect ratios of magnitude.

1. Identify what is represented by each of the following symbols: *X*, *Y*, *N*, *n* and ∑.
2. Perform calculations using summation notation and other mathematical operations following the correct order of operations.

The letter *X* is used to represent scores for a variable. If a second variable is used, *Y* represents its scores. The letter *N* is used as the symbol for the number of scores in a population; *n* is the symbol for a number of scores in a sample.

The Greek letter sigma (Σ) is used to stand for summation. Therefore, the expression ΣX is read “the sum of the scores.” Summation is a mathematical operation (like addition or multiplication) and must be performed in its proper place in the order of operations; summation occurs after parentheses, exponents, and multiplying/dividing have been completed.

**Other Lecture Suggestions**

* 1. Early in the first class, acknowledge that:
* Most students are not there by choice. (No one picked statistics as an elective because it looked like a fun class.)
* Many students have some anxiety about the course.

However, 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:

* 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.
* 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.
1. Give students a list of variables, for example items from a survey (age, gender, education level, income, occupation) and ask students to identify the scale of measurement most likely to be used and whether the variable is discrete or continuous.
2. Describe a non-experimental 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? For example, 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?

 In either case, describe how the study could be made into an experiment by:

* 1. beginning with equivalent groups (random assignment).
	2. manipulating the independent variable (this introduces the ethical question of forcing people to eat a nutritious breakfast).
	3. controlling other variables (the rest of the children’s diet).
1. 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, Σ(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.

 Alternatively, there are frequently social media posts that test knowledge of the order of operations. Google “social media order operations” and click on “images” to see recent ones. Present several to students to review the order of operations. One that claims a certain percentage of people get it wrong will allow an opportunity to discuss the misuse of statistics as well.

1. Invite students to explore how they come into contact with statistics in their everyday lives. Use an article like [Statistics in Everyday Life](http://www.isixsigma.com/community/blogs/statistics-everyday-life/) (<http://www.isixsigma.com/community/blogs/statistics-everyday-life/>) to stimulate discussion. Invite the students to find an article online or in a newspaper about a topic of interest to them and to review how that article uses (or misuses) statistics. Ask them to consider the implications of not understanding statistics and their use.

**Answers to Even-Numbered Problems**

2. The population is the entire group of individuals (or scores) of interest for a particular research study. A sample is a group selected from a population that usually is used to represent the population in a research study. A parameter is a characteristic, usually a numerical value, that describes a population. A statistic is a characteristic, usually numerical, that describes a sample.

4. Sampling error is the naturally occurring difference between a sample and the population from which the sample is obtained. Specifically, the statistics obtained for a sample will be different from the corresponding parameters for the population and the statistics will differ from one sample to another. This is a problem for inferential statistics because any difference found between two treatment conditions may be explained by the treatments but it also may be explained by sampling error.

6. The goal of an experiment is to demonstrate the existence of a cause-and-effect relationship between two variables. To accomplish the goal, an experiment must manipulate an independent variable and control other, extraneous variables.

8. This is not an experiment because no independent variable is manipulated. The researchers are comparing two preexisting groups of individuals.

10. a. The dependent variable is comprehension of the passage, which is measured by the test

 scores.

 b. Knowledge or comprehension is continuous.

 c. ratio scale (zero means no correct answers)

12. a. The independent variable is taking the Tai Chi course versus not taking the course.

 b. Nominal scale

 c. The dependent variable is the amount of arthritis pain experienced.

 d. The amount of pain probably is measured with an interval or a ratio scale.

14. a. An ordinal scale provides information about the direction of difference (greater or less) between two measurements.

 b. An interval scale provides information about the magnitude of the difference between two measurements.

 c. A ratio scale provides information about the ratio of two measurements.

16. Honesty is an attribute or personality characteristic that cannot be observed or measured directly. Shyness could be operationally defined by identifying and observing external behaviors associated with being shy. Or, participants could be given a questionnaire asking how they behave or feel in situations for which shyness might have an influence.

18. a. Σ*X* = 10

 b. Σ*X*2 = 38

 c. Σ*X* + 1 = 11

 d. Σ(*X* + 1) = 14

20. a. Σ*X* = 0

 b. Σ*X*2 = 50

 c. Σ(*X* + 3) = 15

22. a. (Σ*X*)2

 b. Σ*X*2

 c. Σ(*X* – 2)

 d. Σ(*X* – 1)2