CHAPTER 1
Introduction to Statistics

What Are You Getting Into?

WHAT IS STATISTICS?

This book addresses the topic of statistics, but what exactly do we mean by statistics? Statistics is a branch of applied mathematics that involves the collection and interpretation of data and the use of mathematical principles to draw conclusions about the results of our observations. Stated another way, statistics is a field of mathematical study that addresses how data that are collected in a study (such as the number of accidents at a particular intersection) can reveal patterns and how to evaluate those patterns based on the likelihood of that pattern occurring by chance. Patterns that are considered to be unlikely occurrences if chance alone is operating are then said to be occurring for reasons other than chance (such as poor visibility in the intersection or some other variable that we believe might be affecting the outcome of our study).

WHAT IS OUR APPROACH TO STATISTICS?

This book is intended for those of us who use statistics as a tool to understand the data that they collect in their research or to understand the data collected by others in their research. Thus, we will not delve too deeply into the theoretical basis behind the mathematics that we teach, but you will definitely be using some of your skills in mathematics in this course, and it would be a good idea for you to review some basic college mathematics textbooks if you feel you have difficulty recalling things like the proper order that you follow when you have multiple operations in a single problem (e.g., always do what is in the parentheses first). There will be statistical tests that you will learn to conduct by hand. However, while we feel that requiring some calculations to be done by hand is important in teaching the underlying basis of statistical interpretation, this can be done for a limited number of tests and still accomplish the conceptual goals of the course.

In addition, we will emphasize how the steps you take in calculating and evaluating a statistical test by hand compare to interpreting a printout of the same statistical test in two different computer software packages. Because the statistical terminology varies among textbooks and software packages, it is critical that you are trained to understand the conceptual basis to the labels. That will be easier for you if you have been exposed to more than one style of output. Truthfully, very few scientists calculate statistical tests by hand, and when they do, it is only for the simple tests. Thus, we will provide examples in the appropriate chapters that demonstrate the step-by-step calculations done by hand and then show the output from the same data in Microsoft's Excel and SPSS Inc.'s Statistical Package for the Social Sciences (SPSS). We chose these software packages because they are much different from one another but commonly used, and SPSS has a student version. We will highlight the numbers that match the numbers that we calculated by hand, and we will define all of the output since there are more numbers provided by computer software packages than you normally use when you hand calculate the same tests. The additional numbers have proven confusing to students when they get to their laboratory course if they have only been trained in the hand calculation techniques.

Beginning in Part II, you will be introduced to a flowchart that will guide you on how to choose the appropriate statistical test for a given data set and experimental design, and that flowchart will be built up chapter by chapter as you learn more tests. In fact, we feel that choosing the appropriate test is the most important concept in undergraduate statistics because the computer can do the analysis for you only if you know which statistical test is appropriate. Students commonly come to their laboratory courses without

the ability to choose the appropriate test because they have learned statistics one chapter at a time and there was only one statistical test per chapter.

We will also introduce you to the concept of a "general linear model." This is often neglected in an undergraduate course, but it is an important and useful concept. On the basis of our coverage of this model, you will be able to analyze data using the general linear model on a statistics software package (such as SPSS) if you ever need to do so. You will be able to apply this more modern technique to appropriate data sets and will also better understand the mathematical relationships between tests that are commonly performed in the behavioral sciences (independent *t* test, paired *t* test, analysis of variance [ANOVA], linear regression), all of which you will hear about in later chapters.

PRACTICAL DETAILS BEFORE WE CAN MOVE ON

Before we move on into new material in Chapter 2, there are a couple of practical details to cover, including a discussion of rounding and features of a statistical calculator. The details of this material might be altered by your instructor for the specific details of your class, but this information should prove useful for most students.

A Short Discussion of Rounding

The rules about rounding vary a bit based on the field of study. We will follow the rule that is common in the social sciences, which is to round to two decimal places. For example, if you have a number such as 164.0665, we would round that number using rules that are based on the number in the third decimal position.

Rounding Rules

If the number in the third decimal position is > 5, we would round the second decimal position up by one unit. If the number in the third decimal position is < 5, we would retain the number in the second decimal. If the number was exactly 5, then we would look at the next decimal place (fourth) but apply the same "less than or greater than 5" rule.

Example

For the number 164.0665, the number in the third decimal position is a 6, and thus we would round the number in the second decimal position

(also a 6) up one unit (7). Thus, our newly rounded number would be reported as 164.07.

However, the number 164.0652 would be reported as 164.06 since you are forced to go to the fourth decimal number and it is smaller than 5. This technique may differ from the one that you were taught earlier in your education, but following this technique helps avoid bias in rounding that can occur if you round up simply based on the number in the third position being equal to 5.

A Warning About Rounding

Keep in mind that you should not round numbers until you arrive at your final calculation. Rounding as you calculate intermediate numbers can result in the accumulation of small errors.

Buying an Appropriate Calculator

You should avoid buying a calculator that will do more than you need to do in this course. Students often buy "programmable" or "graphing" scientific calculators, and they not only are unnecessary for this course but also often confuse the student who is trying to use them. There are a few symbols that you should look for in a calculator. The symbols sometimes vary by the manufacturer, but we will give you some of the more common variations of the symbols you need to aid you in your purchase. The calculators that we (as instructors) use for this course cost us between U.S.$6 and $8 and were purchased in an office supply store, so you can save yourself some money as well as frustration by following our advice. You do need to find a scientific or statistical calculator and not a business calculator, but you will likely want to stay away from the programmable calculators unless you are extremely proficient with one already.

Symbols You Need on Your Calculator

Σ or a Memory Storage Key, such as M+ or Σ+

ΣX

ΣX^2

$\bar{X}$ or μ

s_n or σ or σ_n

s_{n-1} or s or σ_{n-1}

$\sqrt{}$

X^2

How to Use Your Calculator

It is difficult to give you instructions on how to use your calculator since they vary so much, but there are excellent resources on the Internet to help you with this task. It is very important to take the time to learn how to enter data and to use the memory storage function on your calculator. This will save you significant time and errors on your homework and exams.

SUMMARY

This book will focus on the conceptual and practical aspects of statistics, particularly on how they are used in the behavioral sciences. You should heed our advice to purchase an inexpensive calculator that is easy to use and will aid you in performing well in this course. In the next chapter, we will introduce you to the important terminology used in statistics.