CHAPTER 1

Introduction

1.1 Why Statistics?

An anonymous sage once defined a statistician as "one who collects data and draws confusions." Another declared that members of this tribe occupy themselves by "drawing mathematically precise lines from unwarranted assumptions to foregone conclusions." And then there is the legendary proclamation attributed (by Mark Twain) to the 19th-century British statesman Benjamin Disraeli: "There are three kinds of lies: lies, damned lies, and statistics."

Are such characterizations justified? Clearly we think not! Just as every barrel has its rotten apples, there are statisticians among us for whom these sentiments are quite accurate. But they are the exception, not the rule. While there are endless reasons explaining why statistics is sometimes viewed with skepticism (math anxiety? mistrust of the unfamiliar?), there is no doubt that when properly applied, statistical reasoning serves to illuminate, not obscure. In short, our objective in writing this book is to acquaint you with the proper applications of statistical reasoning. As a result, you will be a more informed and critical patron of the research you read; furthermore, you will be able to conduct basic statistical analyses to explore empirical questions of your own.

Statistics merely formalizes what humans do every day. Indeed, most of the fundamental concepts and procedures we discuss in this book have parallels in everyday life, if somewhat beneath the surface. You may notice that there are people of different ages ("variability") at Eric Clapton concerts. Because Maine summers are generally warm ("average"), you don't bring a down parka when you vacation there. Parents from a certain generation, you observe, tend to drive Volvo station wagons ("association"). You believe that it is highly unlikely ("probability") that your professor will take attendance two days in a row, so you skip class the day after attendance was taken. Having talked for only a few minutes ("sample") with a person you just met, you conclude that you like him ("generalization," "inference"). After getting a disappointing meal at a popular restaurant, you wonder whether it was just an off night for the chef or the place actually has gone downhill ("sampling variability," "statistical significance").

We could go on, but you probably get the point: Whether you are formally crunching numbers or simply going about life, you employ—consciously or not—the fundamental concepts and principles underlying statistical reasoning.

So what does formal statistical reasoning entail? As can be seen from the two-part structure of this book, statistical reasoning has two general branches: descriptive statistics and inferential statistics.

1.2 Descriptive Statistics

Among first-year students who declare a major in education, what proportion are male? Female? Do those proportions differ between elementary education and secondary education students? Upon graduation, how many obtain teaching positions? How many go on to graduate school in education? And what proportion end up doing something unrelated to education? These are examples of questions for which **descriptive statistics** can help to provide a meaningful and convenient way of characterizing and portraying important features of the data.[1] In the examples above, *frequencies* and *proportions* will help to do the job of statistical description.

> The purpose of **descriptive statistics** is to organize and summarize data so that the data are more readily comprehended.

What is the average age of undergraduate students attending American universities for each of the past 10 years? Has it been changing? How much? What about the Graduate Record Examination (GRE) scores of graduate students over the past decade—has that average been changing? One way to show the change is to construct a graph portraying the average age or GRE score for each of the 10 years. These questions illustrate the use of *averages* and *graphs*, additional tools that are helpful for describing data.

We will explore descriptive procedures in later chapters, but for the present let's consider the following situation. Professor Tu, your statistics instructor, has given a test of elementary mathematics on the first day of class. She arranges the test scores in order of magnitude, and she sees that the distance between the highest and lowest scores is not great and that the class average is higher than normal. She is pleased because the general level of preparation seems to be good and the group is not exceedingly diverse in its skills, which should make her teaching job easier. And you are pleased, too, for you learn that your performance is better than that of 90% of the students in your class. This scenario illustrates the use of more tools of descriptive statistics: the *frequency distribution*, which shows the scores in ordered arrangement; the *percentile*, a way to describe the location of a person's score relative to scores of others in a group; and the *range*, which measures the variability of scores.

Because they each pertain to a single variable—age, GRE scores, and so on—the preceding examples involve **univariate** procedures for describing data. But often researchers are interested in describing data involving two characteristics of a person (or object) simultaneously, which call for **bivariate** procedures. For example, if you had information on 25 people concerning how many friends each person has (popularity) and how outgoing each person is (extroversion), you could see whether popularity and extroversion are related. Is popularity greater among people with higher levels of extroversion and, conversely, lower among people lower in extroversion? The *correlation coefficient* is a bivariate statistic that

[1]We are purists with respect to the pronunciation of this important noun ("day-tuh") and its plural status. Regarding the latter, promise us that you will recoil whenever you hear an otherwise informed person utter, "The data is. . . ." Simply put, data *are*.

describes the nature and magnitude of such relationships, and a *scatterplot* is a helpful tool for graphically portraying these relationships.

Regardless of how you approach the task of describing data, never lose sight of the principle underlying the use of descriptive statistics: The purpose is to organize and summarize data so the data are more readily comprehended and communicated. When the question “Should I use statistics?” comes up, ask yourself, “Would the story my data have to tell be clearer if I did?”

1.3 Inferential Statistics

What is the attitude of taxpayers toward, say, the use of federal dollars to support private schools? As you can imagine, pollsters find it impossible to put such questions to *every* taxpayer in this country! Instead, they survey the attitudes of a random **sample** of taxpayers, and from that knowledge they estimate the attitudes of taxpayers as a whole—the **population**. Like any estimate, this outcome is subject to random “error” or **sampling variation**. That is, random samples of the same population don’t yield identical outcomes. Fortunately, if the sample has been chosen properly, it is possible to determine the magnitude of error that is involved.

The second branch of statistical practice, known as **inferential statistics**, provides the basis for answering questions of this kind. These procedures allow one to account for chance error in drawing inferences about a larger group, the population, on the basis of examining only a sample of that group. A central distinction here is that between **statistic** and **parameter**. A statistic is a characteristic of a sample (e.g., the proportion of *polled* taxpayers who favor federal support of private schools), whereas a parameter is a characteristic of a population (the proportion of *all* taxpayers who favor such support). Thus, statistics are used to estimate, or make inferences about, parameters.

> **Inferential statistics** permit conclusions about a population, based on the characteristics of a sample of the population.

Another application of inferential statistics is particularly helpful for evaluating the outcome of an experiment. Does a new drug, Melo, reduce hyperactivity among children? Suppose that you select at random two groups of hyperactive children and prescribe the drug to one group. All children are subsequently observed the following week in their classrooms. From the outcome of this study, you find that, on average, there is less hyperactivity among children receiving the drug.

Now some of this difference between the two groups would be expected even if they were treated alike in all respects, because of chance factors involved in the random selection of groups. As a researcher, the question you face is whether the obtained difference is within the limits of chance sampling variation. If certain assumptions have been met, statistical theory can provide the basis for an answer. If you find that the obtained difference is larger than that can be accounted for by chance alone, you will infer that other factors (the drug being a strong candidate) must be at work to influence hyperactivity.

This application of inferential statistics also is helpful for evaluating the outcome of a correlational study. Returning to the preceding example concerning the relationship between popularity and extroversion, you would appraise the obtained correlation much as you would appraise the obtained difference in the hyperactivity experiment: Is this correlation larger than what would be expected from chance sampling variation alone? If so, then the traits of popularity and extroversion may very well be related in the population.

1.4 The Role of Statistics in Educational Research

Statistics is neither a beginning nor an end. A problem begins with a question rooted in the substance of the matter under study. *Does* Melo reduce hyperactivity? *Is* popularity related to extroversion? Such questions are called **substantive questions**.[2] You formulate the question, which is informed by a careful consideration of relevant theory, associated research, and due regard for the field of educational practice. You then decide on the appropriate methodology for exploring the question empirically—that is, using data.

Once you have diligently crafted the substantive question—and only then—it is now time for statistics to play a part. Let's say your study calls for averages (as in the case of the hyperactivity experiment). You calculate the average for each group and raise a **statistical question**: Are the two averages so different that sampling variation alone cannot account for the difference? Statistical questions differ from substantive questions in that the former are questions about a statistical index—in this case, the average. If, after applying the appropriate statistical procedures, you find that the two averages are so different that it is not reasonable to believe chance alone could account for it, you have made a **statistical conclusion**—a conclusion about the statistical question you raised.

Now back to the substantive question. If certain assumptions have been met and the conditions of the study have been carefully arranged, you may be able to conclude that the drug *does* make a difference, at least within the limits tested in your investigation. This is your final conclusion, and it is a **substantive conclusion**. Although the substantive conclusion derives partly from the statistical conclusion, other factors must be considered. As a researcher, therefore, you must weigh both the statistical conclusion *and* the adequacy of your methodology in arriving at the substantive conclusion.

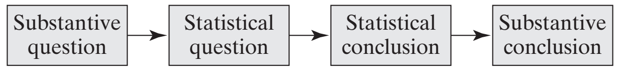
It is important to see that although there is a close relationship between the substantive question and the statistical question, the two are not identical. You will recall that a statistical question always concerns a statistical property of the data (e.g., an average or a correlation). Often, alternative statistical questions can be applied to explore the particular substantive question. For instance, one might ask whether the *proportion* of students with very high levels of hyperactivity differs beyond the limits of chance variation between the two conditions. In this case, the statistical question is about a different statistical index: the proportion rather than the average.

Thus, part of the task of mastering statistics is to learn how to choose among, and sometimes combine, different statistical approaches to a particular substantive question. When designing a study, the consideration of possible statistical analyses

[2]The substantive question also is called the research question.

to be performed should be situated in the course of refining the substantive question and developing a plan for collecting relevant data.

To sum up, the use of statistical procedures is always a middle step; they are a technical means to a substantive end. The argument we have presented can be illustrated as follows:



1.5 *Variables and Their Measurement*

Descriptive and inferential statistics are applied to **variables**.

> A **variable** is a characteristic (of a person, place, or thing) that takes on different values.

Variables in educational research often (but not always) reflect characteristics of *people*—academic achievement, age, leadership style, intelligence, educational attainment, beliefs and attitudes, and self-efficacy, to name a few. Two nonpeople examples of variables are school size and brand of computer software. Although simple, the defining characteristic of a variable—*something that varies*—is important to remember. A "variable" that doesn't vary sufficiently, as you will see later, can sabotage your statistical analysis.[3]

Statistical analysis is not possible without numbers, and there cannot be numbers without **measurement**.

> **Measurement** is the process of assigning numbers to the characteristics you want to study.

For example, "20 years" may be the measurement for the characteristic, *age*, for a particular person; "115" may be that person's measurement for *intelligence*; on a scale of 1 to 5, "3" may be the *sociability* measurement for this person; and because this hypothetical soul is female, perhaps she arbitrarily is assigned a value of "2" for *sex* (males being assigned "1").

But numbers can be deceptive. Even though these four characteristics—age, intelligence, sociability, and sex—all have been expressed in numerical form, the numbers differ considerably in their underlying properties. Consequently, these numbers also differ in how they should be interpreted and treated. We now turn to a more detailed consideration of a variable's properties and the corresponding implications for interpretation and treatment.

[3]If this statement perplexes you, think through the difficulty of determining the relationship between, say, "school size" and "academic achievement" if *all* of the schools in your sample were of an identical size. How could you possibly know whether there is a relationship (correlation) between academic achievement and school size? True, this may be an absurd scenario. After all, who would ask such a question if all schools were the same size?! But the problem of insufficient variability is far more subtle in practice, as you will see in Chapter 7.

Qualitative Versus Quantitative Variables

Values of **qualitative variables** (also known as **categorical variables**) differ in *kind* rather than in *amount*. Sex is a good example. Although males and females clearly are different in reproductive function (a qualitative distinction), it makes no sense to claim one group is either "less than" or "greater than" the other in this regard (a quantitative distinction).[4] And this is true even if the arbitrary measurements suggest otherwise! Other examples of qualitative variables are college major, marital status, political affiliation, county residence, and ethnicity.

In contrast, the numbers assigned to **quantitative variables** represent differing *quantities* of the characteristic. Age, intelligence, and sociability, which you saw above, are examples of quantitative variables: A 40-year-old is "older than" a 10-year-old; an IQ of 120 suggests "more intelligence" than an IQ of 90; and a child with a sociability rating of 5 presumably is more sociable than the child assigned a 4. Thus, the values of a quantitative variable differ in amount. As you will see shortly, however, the properties of quantitative variables can differ greatly.

Scales of Measurement

In 1946, Harvard psychologist S. S. Stevens wrote a seminal article on **scales of measurement**, in which he introduced a more elaborate scheme for classifying variables. Although there is considerable debate regarding the implications of his typology for statistical analysis (e.g., see Gaito, 1980; Stine, 1989), Stevens nonetheless provided a helpful framework for considering the nature of one's data. A variable, Stevens argued, rests on one of four scales: **nominal**, **ordinal**, **interval**, or **ratio**.

Nominal scales Values on a nominal scale merely "name" the category to which the object under study belongs. As such, interpretations must be limited to statements of kind rather than amount. (A qualitative variable thus represents a nominal scale.) Take ethnicity, for example, which a researcher may have coded 1 = Italian, 2 = Irish, 3 = Asian, 4 = Hispanic, 5 = African American, and 6 = Other.[5] It would be perfectly appropriate to conclude that, say, a person assigned "1" (Italian, we trust) is different from the person assigned "4" (Hispanic), but you cannot demand more of these data. For example, you could not claim that because $3 < 5$, Asian is "less than" African American; or that an Italian, when added to an Asian, begets an Hispanic (because $1 + 3 = 4$). The numbers don't mind, of course, but it still makes no sense. The moral throughout this discussion is the same: One should remain forever mindful of the variable's underlying scale of measurement and the kinds of interpretations and operations that are sensible for that scale.

Ordinal scales Unlike nominal scales, values on an ordinal scale can be "ordered" to reflect differing degrees or amounts of the characteristic under study. For example, rank ordering students based on when they completed an in-class exam would reflect an ordinal scale, as would ranking runners according to when

[4]Although males and females, on average, do differ *in amount* on any number of variables (e.g., height, strength, annual income), the scale in question is no longer sex. Rather, it is the scale of the other variable on which males and females are observed to differ.

[5]Each individual must fall into only one category (i.e., the categories are *mutually exclusive*), and the five categories must represent all ethnicities included among the study's participants (i.e., the categories are *exhaustive*).

they crossed the finish line. You know that the person with the rank of 1 finished the exam sooner, or the race faster, than individuals receiving higher ranks.[6] But there is a limitation to this additional information: The only relation implied by ordinal values is "greater than" or "less than." One cannot say *how much* sooner the first student completed the exam compared to the third student, or that the difference in completion time between these two students is the same as that between the third and fourth students, or that the second-ranked student completed the exam in half the time of the fourth-ranked student. Ordinal information simply does not permit such interpretations.

Although rank order is the classic example of an ordinal scale, other examples frequently surface in educational research. Percentile ranks, which we take up in Chapter 2, fall on an ordinal scale: They express a person's performance relative to the performance of others (and little more). Likert-type items, which many educational researchers use for measuring attitudes, beliefs, and opinions (e.g., 1 = strongly disagree, 2 = disagree, and so on), are another example. Socioeconomic status, reflecting such factors as income, education, and occupation, often is expressed as a set of ordered categories (e.g., 1 = lower class, 2 = middle class, 3 = upper class) and, thus, qualifies as an ordinal scale as well.

Interval scales Values on an interval scale overcome the basic limitation of the ordinal scale by having "equal intervals." The 2-point difference between, say, 3 and 5 on an interval scale is the same—in terms of the underlying characteristic being measured—as the difference between 7 and 9 or 24 and 26. Consider an ordinary Celsius thermometer: A drop in temperature from 30°C to 10°C is equivalent to a drop from 50°C to 30°C.

The limitation of an interval scale, however, can be found in its **arbitrary zero**. In the case of the Celsius thermometer, for example, 0°C is arbitrarily set at the point at which water freezes (at sea level, no less). In contrast, the absence of heat (the temperature at which molecular activity ceases) is roughly –273°C. As a result, you could not claim that a 30°C day is three times as warm as a 10°C day. This would be the same as saying that column A in Figure 1.1 is three times FIG1

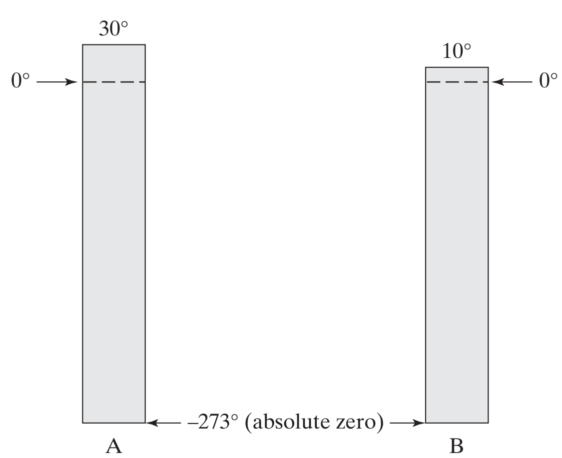


Figure 1.1 Comparison of 30° and 10° with the absolute zero on the Celsius scale.

S
N
L

[6]Although perhaps counterintuitive, the convention is to reserve *low* ranks (1, 2, etc.) for good performance (e.g., high scores, few errors, fast times).

as tall as column B. Statements involving ratios, like the preceding one, cannot be made from interval data.

What are examples of interval scales in educational research? Researchers typically regard composite measures of achievement, aptitude, personality, and attitude as interval scales. Although there is some debate as to whether such measures yield truly interval data, many researchers (ourselves included) are comfortable with the assumption that they do.

Ratio scales The final scale of measurement is the ratio scale. As you may suspect, it has the features of an interval scale *and* it permits ratio statements. This is because a ratio scale has an **absolute zero**. "Zero" weight, for example, represents an unequivocal absence of the characteristic being measured: *no* weight. Zip, nada, nothing. Consequently, you can say that a 230-pound linebacker weighs twice as much as a 115-pound jockey, a 30-year-old is three times the age of a 10-year-old, and the 38-foot sailboat *Adagio* is half the length of 76-foot *White Wings*—for weight, age, and length are all ratio scales.

In addition to physical measures (e.g., weight, height, distance, elapsed time), variables derived from *counting* also fall on a ratio scale. Examples include the number of errors a student makes on a reading comprehension task; the number of friends one reports having; the number of verbal reprimands a high school teacher issues during a lesson; or the number of students in a class, school, or district.

As with any scale, one must be careful when interpreting ratio scale data. Consider two vocabulary test scores of 10 and 20 (words correct). Does 20 reflect twice the performance of 10? It does if one's interpretation is limited to performance on this particular test ("You knew twice as many words on this list as I did"). However, it would be unjustifiable to conclude that the student scoring 20 has twice the *vocabulary* as the student scoring 10. Why? Because "0" on this test does not represent an absence of vocabulary; rather, it represents an absence of knowledge of the specific words on this test. Again, proper interpretation is critical with any measurement scale.

1.6 Some Tips on Studying Statistics

Is statistics a hard subject? It is and it isn't. Learning the "how" of statistics requires attention, care, and arithmetic accuracy, but it is not particularly difficult. Learning the "why" of statistics varies over a somewhat wider range of difficulty.

What is the expected reading rate for a book about statistics? Rate of reading and comprehension differ from person to person, of course, and a four-page assignment in mathematics may require more time than a four-page assignment in, say, history. Certainly, you should not expect to read a statistics text like a novel, or even like the usual history text. Some parts, like this chapter, will go faster; but others will require more concentration and several readings. In short, do not feel cognitively challenged or grow impatient if you can't race through a chapter and, instead, find that you need time for absorption and reflection. The formal logic of statistical inference, for example, is a new way of thinking for most people and requires some getting used to. Its newness can create difficulties for those who are not willing to slow down. As one of us was constantly reminded by his father, "*Festina lente!*"[7]

Many students expect difficulty in the area of mathematics. Ordinary arithmetic and some familiarity with the nature of equations are needed. Being able to see "what goes on" in an equation—to peek under the mathematical hood, so to speak—is necessary to understand what affects the statistic being calculated, and in what way. Such understanding also is helpful for spotting implausible results, which allows you to catch calculation errors when they first occur (rather than in an exam). Appendix A is especially addressed to those who feel that their mathematics lies in the too-distant past to assure a sense of security. It contains a review of elementary mathematics of special relevance for study of this book. Not all these understandings are required at once, so there will be time to brush up in advance of need.

Questions and problems are included at the end of each chapter. You should work enough of these to feel comfortable with the material. They have been designed to give practice in how-to-do-it, in the exercise of critical evaluation, in development of the link between real problems and methodological approach, and in comprehension of statistical relationships. There is merit in giving some consideration to all questions and problems, even though your instructor may formally assign fewer of them.

A word also should be said about the cumulative nature of a course in elementary statistics: What is learned in earlier stages becomes the foundation for what follows. Consequently, it is most important to keep up. If you have difficulty at some point, seek assistance from your instructor. Don't delay. Those who think matters may clear up if they wait may be right, but the risk is greater here—considerably so—than in courses covering material that is less interdependent. It can be like attempting to climb a ladder with some rungs missing, or to understand an analogy when you don't know the meaning of all the words. Cramming, never very successful, is least so in statistics. Success in studying statistics depends on regular work, and, if this is done, relatively little is needed in the way of review before examination time.

Finally, always try to "see the big picture." First, this pays off in computation. Look at the result of your calculation. *Does it make sense?* Be suspicious if you find the average to be 53 but most of the numbers are in the 60s and 70s. And dismiss outright any result that defies plausibility, such as a statistic that falls beyond its possible limits. Remember, the eyeball is the statistician's most powerful tool. Second, because of the ladderlike nature of statistics, also try to relate what you are currently studying to concepts, principles, and techniques you learned earlier. Search for connections—they are there. When this kind of effort is made, you will find that statistics is less a collection of disparate techniques and more a concerted course of study. Happily, you also will find that it is easier to master!

[7]"Make haste slowly!"

Exercises

Identify, Define, or Explain

Terms and Concepts

descriptive statistics	substantive conclusion
univariate	variable
bivariate	measurement
sample	qualitative variable (or categorical variable)
population	quantitative variable
sampling variation	scales of measurement
inferential statistics	nominal scale
statistic	ordinal scale
parameter	interval scale
substantive question	ratio scale
statistical question	arbitrary zero
statistical conclusion	absolute zero

Questions and Problems

Note: Answers to starred (*) items are given in Appendix B.

*1. Indicate which scale of measurement each of the following variables reflects:
   - **(a)** the distance one can throw a shotput
   - **(b)** urbanicity (where 1 = urban, 2 = suburban, 3 = rural)
   - **(c)** school locker numbers
   - **(d)** SAT score
   - **(e)** type of extracurricular activity (e.g., debate team, field hockey, dance)
   - **(f)** university ranking (in terms of library holdings)
   - **(g)** class size
   - **(h)** religious affiliation (1 = Protestant, 2 = Catholic, 3 = Jewish, etc.)
   - **(i)** restaurant rating (* to ****)
   - **(j)** astrological sign
   - **(k)** miles per gallon

2. Which of the variables from Problem 1 are qualitative variables and which are quantitative variables?

3. For the three questions that follow, illustrate your reasoning with a variable from the list in Problem 1.
   - **(a)** Can a ratio variable be reduced to an ordinal variable?
   - **(b)** Can an ordinal variable be promoted to a ratio variable?
   - **(c)** Can an ordinal variable be reduced to a nominal variable?

*4. Round the following numbers as specified (review Appendix A.7 if necessary):
   - **(a)** to the nearest whole number: 8.545, −43.2, 123.01, .095
   - **(b)** to the nearest tenth: 27.33, 1.9288, −.38, 4.9746
   - **(c)** to the nearest hundredth: −31.519, 76.0048, .82951, 40.7442

S
N
L

5. In his travels, one of the authors once came upon a backroad sign announcing that a small town was just around the corner. The sign included the town's name, along with these facts:

Population	562
Feet above sea level	2150
Established	1951
TOTAL	4663

Drawing on what you have learned in this chapter, evaluate the meaning of "4663."