1 Statistics in the Research Process

1.1 Introduction

This book is about how to use and apply statistics. National and international newspapers, magazines, books, and scientific journals are some of the many documents published where some sort of statistical information is reported. Almost every media source requires that we have some statistical knowledge to understand and make the right conclusion about the material being presented.

In politics, elections include claims and counterclaims by political opponents. Incumbent candidates may say that during their terms, the city's crime rate decreased by 20 percent, for example. Candidates want voters to think they were responsible for the crime reduction. Opponents will cite their own statistical information that may show that crime actually increased by 30 percent. Disputes follow where each candidate's statistics are closely scrutinized by citizens and political commentators. Quite often, these different opponents are citing accurate crime figures; however, the types of crimes included in their statistics may be different and coming from different points in time. Thus, the results seem contradictory, but they are not. Each side uses statistical information to its own advantage in an attempt to earn support from voters. The public is understandably confused by these contradictory (and oftentimes manipulated) results, which are technically correct in the statistical universe.

In health-related affairs, governmental agencies around the world screen, evaluate, and limit the distribution of drugs and other products for public consumption. Health-related recommendations are accompanied by statistical evidence and findings. Statistical results from various scientific investigations are responsible for most agency assertions and decisions.

These examples are only a few of many illustrating the crucial role of statistics in influencing social policies, as well as in describing and predicting social behavior in general. Statistics are a pervasive and vital part of our lives. Our society is complex, and it is important for us to know what statistics are and how statistical information can benefit us. Knowledge of statistics assists us in understanding the literature we read, particularly the professional journals published in our respective academic areas of criminology, cognitive sciences, and other social sciences. Statistical reports are typically reported in the results section of a scientific article. The results in conjunction with the methodological and analytic sections of the article are the most crucial and objective parts of the article. It is also important to understand the various limitations of statistics. Sometimes statistical information is used as the basis for largely unsupported claims about research findings. It is important to understand statistics enough to know the difference between responsible (ethical) and irresponsible (unethical) claims that have allegedly been supported by statistical results.

1.2 What Are Statistics?

Statistics is polysemic in the sense that it has many meanings. Generally, it refers to the body of mathematical procedures used to assemble, describe, and summarize (**descriptive statistics**) and tools used to infer facts from numerical data (**inferential statistics**). Information we collect about people's attitudes on altruistic behavior, prejudice, stereotypes, incarceration, and mental health, for example, is converted to numerical data. These numbers are rearranged and analyzed in different ways. We study these numerical patterns. Based on our studies of these numbers and how they are distributed, we learn much about the phenomena we are studying. We ultimately make inferences or causal statements about our findings. If we are interested in studying how social stereotypes influence our perceptions about other ethnicities, our procedures to measure them must be methodologically sound and objective. Is our experiment so well controlled that we can conclude that social stereotypes are culturally ingrained? Statistics, as an inference tool, will help us make decisions about our results. Ultimately, however, it is not what we get (i.e., statistically sound analyses) but how we arrived at our final results (i.e., sound methodology) that determines the accuracy and preciseness of our findings and interpretations.

Statistics also refers to numerical, graphic, and tabular descriptions of collected data. Descriptive statistics, such as the mode, median, mean, range, and standard deviation, are discussed in Chapters 3 and 4. Correlation (Chapter 10) is also considered a descriptive statistic. These are different numerical quantities that describe certain characteristics of the behavior we investigate. Researchers include graphs, tables, and other graphical representations such as those illustrated and discussed in Chapter 2. For example, we might read a research article that has many charts, figures, tables, graphs, and formulas. The article is full of statistics. Numerical and graphic data are presented to summarize what the researchers have found.

Statistics also involves the application of tests and procedures that can be used to analyze and make general inferences about collected data. Social and behavioral scientists have at their disposal a wide array of tests and procedures to help them make objective decisions about their research findings. These statistical procedures compose a common core of tools to help us answer our research questions.

Statistics refers to the characteristics of samples of people about whom we have collected information. This meaning of statistics is most relevant when we distinguish between **populations** and **samples**. A population is the set of all individuals of interest in a study. Populations have certain characteristics, such as average IQs, educational levels, and attitudinal traits. Ordinarily, we do not investigate entire populations of people. Rather, we study (representative) samples or subsets of the population of interest of these people and make generalizations about the populations they represent. For example, we might be interested in investigating attitudes toward environmental protection among 3,000 first-year university students. All 3,000 students would be our population of interest. It would be very labor intensive, however, to survey or interview each and every one of these students. So, what we need to do is choose a sample of 100 students at random. Results from those 100 students would be generalized to the entire population of first-year university students. (In Chapter 7, Section 7.2.2, you'll learn how to determine the exact number of participants needed in a study to obtain a predicted effect, if, in fact, the effect is real.) The characteristics of these samples are also called **statistics** to distinguish

them from population characteristics. These latter characteristics that describe populations are called **parameters**. Traditionally, we use Greek letters to represent parameters and Roman letters to represent statistics or estimates. Thus, the Greek lowercase μ (mu) represents the parameter for the population mean and $\bar{X}$ (pronounced *x-bar*) the estimate (statistic) or sample mean of the parameter μ. Most often, researchers study sample statistics to learn about population parameters. Inferences are made about populations based on sample information. This is part of inferential statistics, which will be examined in later chapters.

1.3 Theory, Research, and Statistics

Theory and research are invariably linked. Theory depends on research to verify it, while research needs theory to guide it. Theory explains events, while research tests these explanations in different ways. Sound research is guided by **theory** or theoretical schemes. To illustrate, consider Karl Popper, a philosopher of science who, in an attempt to demonstrate the importance of theory, asked a group of students to look. *Look*, *look*, *look*! he exclaimed. Understandingly, not knowing where to look, students became annoyed when they kept being asked to look and look. The point was that they didn't know where to look because they didn't know what to look for; indeed, theory guides our research and tells us where, and what, to look for. A theory is an integrated body of assumptions, propositions, and definitions related in ways that explain and predict relationships between two or more variables. Theories are designed to explain research findings. Researchers pose research questions they want to answer. Then they provide one or more explanations as motivation for their proposed studies. These rationales are called **frames of reference**, or ways of looking at research problems. They are usually based on the researchers' experiences. These explanations are subsequently tested by analyzing collected data and applying statistical tests.

Academic journals in the social and behavioral sciences contain articles with accompanying theories about relationships between different phenomena studied. These theoretical schemes explain to readers the logical interrelatedness of different variables. We read these theoretical accounts and explanations for why certain variables are or should be related, and then we examine the findings presented. The information collected, analyzed, and interpreted by investigators are called research findings. Statistical tests and procedures provide us with an array of objective standards that yield more consistent interpretations of collected information. It is more difficult to challenge the objectivity of numerical standards compared with one's emotional beliefs or personal opinions.

1.4 The Circularity of the Research Process

Many components make up the research process. Each part of this process can be viewed as a single link in a chain of events. As Figure 1.1 illustrates, a consistent circularity is associated with the research process.

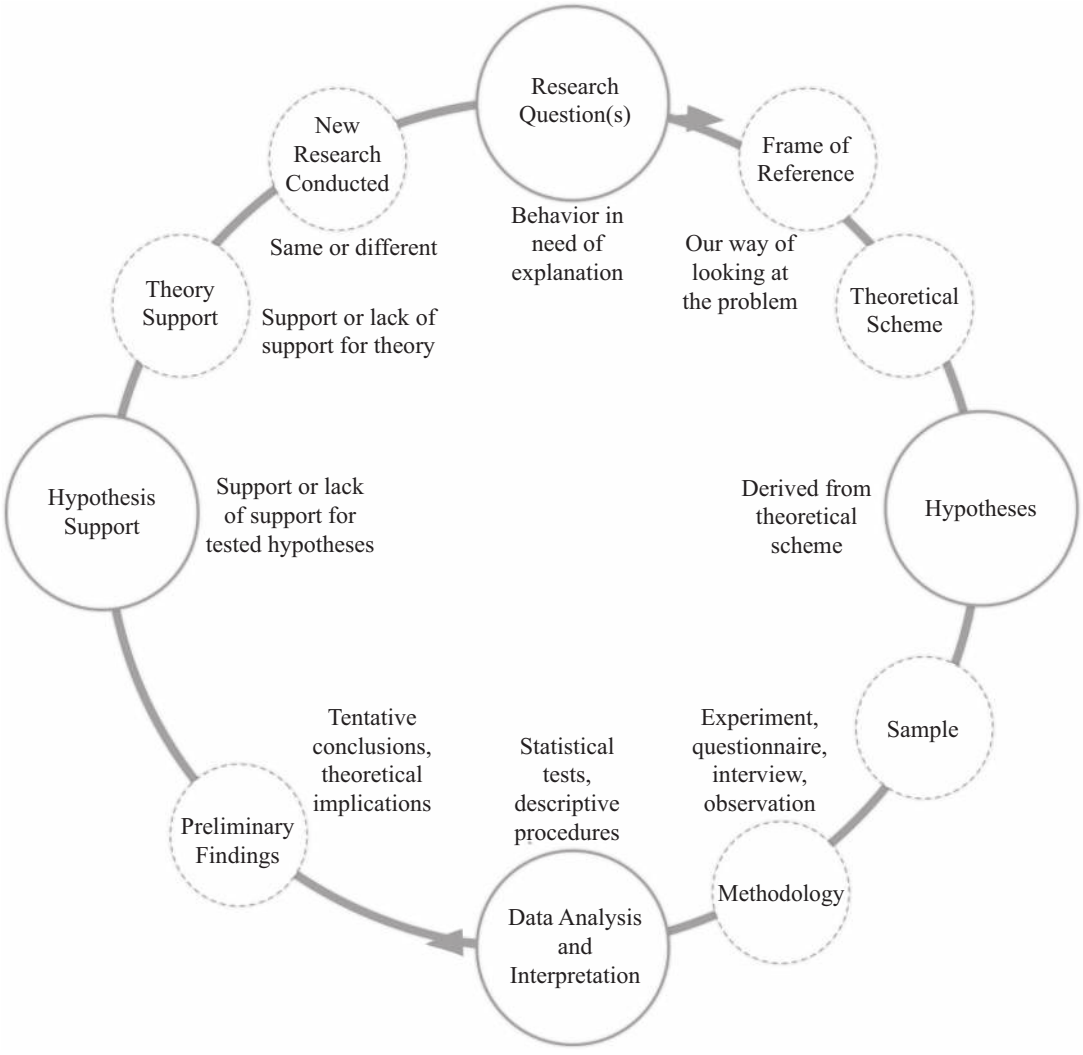

Figure 1.1 The circularity of the research process.

The research process begins with a research question in need of an explanation, such as the issue of perceived power and prison brutality. An explanation for the problem is given, such that guard-on-inmate brutality might be due to situational or environmental factors. This explanation is a frame of reference and an indicator of how and why we believe prison brutality occurs. A theoretical scheme is devised to explain why and how prison guards might be situationally influenced by roles and norms (e.g., finding oneself in a position of power) that lead to unjustified aggression toward inmates. Our theoretical scheme explains and predicts relationships between our research problem and our explanation for it. It also generates hypotheses that can be tested. One way of looking at hypotheses is that they are statements of theory in a testable form. In the perception of the power–prison brutality relationship, for instance, hypotheses are tentative statements about the environmental or situational factors (i.e., roles and norms) and guard brutality as deduced

from our theory. These statements might be tested by gathering data from experiments in simulated prison environments in which, for example, ordinary undergraduate students, with no known mental health issues, are randomly assigned to a prison guard or a prison inmate group. Rates of aggressive behavior, as they exercise control and order, and abuse of power emanating from prison guards, in addition to inmate-specific behaviors (i.e., obedience, fear, and shame) displayed by the inmate group, are recorded and analyzed. As a part of our data analysis, we may use statistical procedures to describe and make decisions about the two experimental groups. We may also apply statistical tests to determine further differences between the two groups in relation to other variables, such as compassionate versus callous and authoritative personality traits.

The chain of events in Figure 1.1 shows that each part of the research process involves doing something to find answers for our original research question(s). Statistics and statistical tests, as well as descriptive procedures, have been placed in the chain involving data analysis and interpretation. The role of statistical procedures in the research process is to provide numerical descriptions of our collected data as well as objective confirmation of hypotheses we test. Hypotheses are statements derived from theories of events that are empirically tested. We test hypotheses in many different ways. Applying particular statistical tests and procedures to our hypotheses is one method for determining whether our hypotheses have predictive validity.

The primary purpose of statistical tests is to help us describe and make decisions about what we have found based on the data we have collected. Statistical results either provide support for or fail to provide support for our theories of why certain events occur and our explanations of these events. Regardless of whether statistical tests provide support for our theories, most researchers conduct additional investigations or replications, often of the same phenomena.

Thus, the work of social scientists never ends. The research process in the social and behavioral sciences is continual. One reason is that no research project is perfect. It is quite common to see in published research phrases such as *more research is needed.* This is because each research project doesn't prove anything. There are no absolutes, since all research conclusions are probabilistic in nature. Rather, more information is added to clarify what is being investigated. It takes a lot of research findings and replications on any given subject to determine which potential variables are most likely (e.g., environmental factors) and which ones are least likely (e.g., personality variables) to cause otherwise law-abiding citizens to commit acts of brutality under the auspices of legal authority.

1.5 Variables

Variables are any phenomena or characteristics that can assume different values. Prison brutality, perceived power, prison guard, and inmate are all variables. The research questions we posed in Section 1.4 involved variables. The explanations we use to account for research questions are also variables. Theories we create explain and predict relations between two or more variables. Thus, we need to know about variables and how they are involved in the research process. Variables studied by investigators include language, sex, socioeconomic status, age, ethnicity, education, political affiliation, type of crime, prior record, job satisfaction, stress, burnout, crime rates, and peer group influence. It is apparent that everything we choose to study is a variable.

1.5.1 Discrete Variables

Variables may be either discrete or continuous. **Discrete variables** have a set of fixed values. For instance, political affiliation is a discrete variable, because political affiliation is only classifiable into a set of fixed categories, such as Democrat, Republican, independent, or some other political party. This means that once you register as a Republican, for example, you cannot be registered as a Democrat. More importantly, you can only vote as a Republican or a Democrat, not as both! Religion is also a discrete variable, with subcategories fixed according to whether a person is Jewish, Catholic, Protestant, or of some other religious faith. Race is also discrete, divided according to categories (e.g., White, Black or African American, American Indian or Alaska Native, Asian, Native Hawaiian or other Pacific Islander).

For data analysis purposes, numbers are assigned to these discrete categories, for instance, Catholic = 1, Jewish = 2, and Muslim = 3. The numbers themselves merely identify different discrete levels or subclasses on these variables. These numbers lack any numerical value or property; they are simply identifiers or categories. Such variables may also be **dichotomous variables**. Dichotomous variables are naturally divisible into discrete categories, such as voted in last election versus didn't vote, felony versus misdemeanor, or 0 versus 1.

1.5.2 Continuous Variables

Continuous variables are any phenomena that can have an infinite or unlimited number of values. Age is a continuous variable, because it can be infinitely divided into years, months, days, hours, minutes, seconds, and so on. In this case, a "1" has a numerical value, and it can range from 1.1 to 1.99. Income is another variable that can have unlimited subdivisions or subcategorizations. Attitudinal measures are continuous variables, since they can assume an infinite number of values from low to high (e.g., low anxiety = 1 to high anxiety = 10; low job satisfaction = 1 to high job satisfaction = 10), where lower scores (e.g., 1.1, 1.2, 2.1, 2.2, 3.0, 3.1) describe lower levels of anxiety and higher scores (e.g., 8.1, 8.2, 9.01, 9.06, 10.1, 10.2) describe higher levels of anxiety. Usually, these degrees of variation for different attitudes are displayed as **raw scores**. Raw scores are generated from measurements or scales we create.

As the result of measuring job satisfaction among college professors, for example, we might assign numbers to different levels of job satisfaction according to the raw scores we have assigned from a job satisfaction scale, as follows: low job satisfaction = 1; moderately low job satisfaction = 2; moderately high job satisfaction = 3; and high job satisfaction = 4. In this and other instances with other variables, raw scores are usually grouped into different categories from low to high, and numbers are assigned. These categories are considered discrete categories, even though they are based on continuous properties underlying their measurement. In effect, any continuous variable can be divided into a sequence of any number of discrete categories for data analysis purposes or even portrayed as dichotomous variables. Age, for example, is a continuous variable that can be converted into a dichotomous variable of young versus old. In many instances, the scores themselves are analyzed directly. If two or more groups of persons are being compared, their scores from various scales may be compared in various ways to determine differences between these groups of persons or samples.

The importance of whether variables are discrete or continuous lies in that each statistical test or technique is accompanied by various assumptions that must be satisfied to apply such tests correctly. One assumption of each statistical test is that the variable being analyzed or tested should be either discrete or continuous. For instance, if we intended to use a particular statistical test that assumed the data should be continuous, and if the data we were analyzing were only discrete, this fact may dissuade us from using the statistical test we originally chose. We might wish to choose another statistical test where this particular assumption is satisfied. The good thing is that there are a wide variety of statistical tests. In many cases, two or more tests exist that perform identical functions. But each has different assumptions that must be met to be applied correctly. One test may assume that the variable to be analyzed must be discrete, while another test may require that the variable analyzed must be continuous. This is only one assumption of several, however. Most statistical tests have at least three or more assumptions underlying their correct application.

1.6 Assumptions Underlying Statistical Procedures

All statistics have assumptions associated with their proper application for describing data, making decisions about hypothesis tests, and/or making inferences about populations from sample characteristics. Thus, for each statistic presented in this book, we state these assumptions.

Researchers make at least three main assumptions when using statistical tests or techniques for descriptive or inferential purposes. These include (1) randomization, (2) level of measurement, and (3) sample size. There are several other assumptions as well, depending on the statistical test or technique presented. But understanding their relevance and importance is based on principles that will be discussed in later chapters, where appropriate.

1.6.1 Randomization

Randomization or random selection refers to how samples are obtained from populations. In statistics, people are often referred to as **elements**. Thus, a sample of persons may be a sample consisting of forty elements. Random selection is defined as drawing elements from a population in such a way that each element has an equal and independent chance of being included in the subsequent sample. An equal chance of being drawn, or **equality of draw**, means that all elements in the population have the same chance of being included. This means that we must be able to identify all elements in the population to give them an equal chance for inclusion. Therefore, if there are 200 population elements, each element has 1/200th of a chance of being drawn. **Independence of draw** or having an independent chance of being included means that drawing one or more elements for inclusion in a sample will not affect the chances of the remaining population elements being included. Samples drawn under these conditions are called **random samples**. Random assignment, in contrast, refers to how people are assigned to different groups, if the research involves experimental and control groups. Researchers nowadays use random-number generators to select samples from designated populations. Any sample drawn from a **table of random numbers** or from a computer-determined draw is considered a random sample. Most statistical tests and procedures assume randomization.

1.6.2 Levels of Measurement

Level of measurement refers to the meanings associated with numbers we assign to collected data. The level of measurement determines which statistical procedures can and cannot be used with the collected data. Numbers may be assigned to attitudinal scores, political affiliation, age, and other variables. Quantifying one's collected data makes the data easier to analyze. For job satisfaction, we may assign numbers to persons who share similar raw scores derived from a job satisfaction scale, or high job satisfaction = 1; moderately high job satisfaction = 2; moderately low job satisfaction = 3; and low job satisfaction = 4. Categories of age may also be assigned numbers: 20 or younger = 1; 21–25 = 2; 26–29 = 3; 30 or older = 4.

Each of these assignments of numbers to different variable subclasses results in a particular level of measurement. When we assign numbers to discrete variables such as political affiliation, type of crime, or religious affiliation, the numbers themselves serve only to differentiate one from another. Republican = 1 and Democrat = 2 doesn't mean, for instance, that a "1" is higher or lower than a "2." It just means the numbers are different, because they are intended to stand for different discrete subclasses. In the case of job satisfaction, a "1" is higher than a "2," which is higher than a "3," which is higher than a "4." Persons assigned a "1" on job satisfaction have "high job satisfaction," while those assigned a "4" have "low job satisfaction." The meanings of numbers, therefore, are dependent on whatever they stand for.

For instance, in a Major League Soccer game, if FC Dallas plays three times and scores two, five, and four goals, we would readily sum the goals and divide by 3, the number of games, to get the team's arithmetic goal average. In this case, $(2+5+4)/3 = 11/3 = 3.7$. The team's goal average for these three games would be 3.7. This average means something. But what about Republican = 1 and Democrat = 2? What if we have a sample with twenty Republicans and twenty Democrats? If we summed the twenty "1" (20×1) values and the twenty "2" (20×2) values, we would have 20 and 40. If we sum these values and divide by the total number of Republicans and Democrats, we would have $(20+40)/40$ or $60/40 = 1.5$. What does this average mean? Does it make sense to say that the average political affiliation is 1.5? Averaging political affiliation is a meaningless calculation. The point is that numbers take on different meanings depending on the scale or level of measurement.

The four levels of measurement, from the lowest to the highest, are (1) the nominal level, (2) the ordinal level, (3) the interval level, and (4) the ratio level. The **nominal level of measurement** is the assignment of names (categories) or numbers to subclasses of variables simply to differentiate between them. Catholic = 1, Protestant = 2, and Jewish = 3 is an assignment of numbers to different religious faiths. It is not expected that we will average these numbers. The importance of these numerical assignments is simply to give us a tally of the numbers of Catholics, Protestants, and Jews in a given sample. Averaging is both impermissible and meaningless. Nominal measurements are discrete and qualitative variables.

Ranking job satisfaction or any other attitude from the most important = 1 to the least important = 4 yields scores according to the **ordinal** or **rank-order level of measurement**. We can say not only that 1s, 2s, 3s, and 4s are different from one another but that they are higher or lower than each other. The different scores represent graduated or ranked categories where certain persons have higher or lower levels of given attitudes compared with other persons. Assigning numbers to data measured according to an ordinal scale permits researchers to make *greater*

than or *less than* distinctions between different scores; differences and direction of differences between scores can be also determined. However, we cannot know or determine actual distances or differences between scores on such scales. We can surmise that a score of "1" is better than a "2," but it would be difficult to know the magnitude of the difference. Typically, ordinal-level measurements are discrete numeric and quantitative variables. However, they can also be non-numeric and qualitative when used as verbal descriptors, such as *large*, *medium*, and *small*.

Ordinal-level measurements should not be averaged, because there must be equal spacing between scores for averaging to be meaningful. But in the social sciences and behavioral sciences, ordinal variables are averaged frequently. Whenever averaging is done for scores measured according to an ordinal scale, caution should be exercised when interpreting the results of such data analyses and statistical applications.

The next highest level of measurement is the **interval level of measurement**. Data measured according to the interval level of measurement are also assigned numbers. These numbers permit some differentiations between values. Furthermore, these numbers permit determinations of *greater than* or *less than*. Additionally, these numbers have equal spacing along an intensity continuum, and researchers may say that there are equal distances between units. Therefore, interval variables are numeric, quantitative, and continuous. Consider the following continuum in displaying various ages:

5 10 15 20 25 30 35 40 45 50 55 60 65 70

This hypothetical age continuum is graduated according to 5-point intervals. These intervals are considered to be of equal distance from each other. For instance, the distance between 5 and 10 is identical to the distance between 50 and 55. The distance between 10 and 30 would be equal to the distance between 50 and 70. Other examples of interval levels of measurement include temperature (in Fahrenheit) and grade point average (GPA). So, for temperature, the difference between 30° and 40° would be the same as the difference between 80° and 90°. This equal spacing of interval scales is desirable, because it permits us to use statistical procedures and other data analysis techniques that involve arithmetic operations, such as averaging, and basic arithmetic operations, such as the computation of square roots.

The **ratio level of measurement** is identical to the interval level of measurement, with one exception: The ratio level of measurement has an absolute zero. Zero, in this case, stands for the *absence of something*. Income is a ratio-level variable, since income may be measured according to a scale having an absolute zero. Persons can have no money. Where an absolute zero is assumed, values may be proportionately related to other values. Therefore, $50 is to $100 as $10,000 is to $20,000. Interval scales lack an absolute zero, and therefore, ratio statements are not permissible with such scales. Thus, temperatures in Fahrenheit can be below zero. A temperature of 0°, in this case, does not reflect a total absence of temperature. Other ratio-level examples include time and weight. (How about age? In Korean culture, for example, a newborn infant is considered to already be one or two years old.)

While income is commonly measurable according to a ratio scale, it is treated as though it were at the interval level of measurement. Since income has ratio-level properties, it also embraces properties of all other lower levels of measurement, including the interval level, the ordinal level, and the nominal level. Various statistical measures and techniques are associated with different levels of measurement. Therefore, it is important to know how variables are

Table 1.1 Levels of measurement and measures of central tendency

Level of measurement	Examples	Statistic
Nominal	F, M	Mode
Ordinal	S, M, L	Mode, median
Interval	Temperature	Mode, median, mean
Ratio	Age, weight	Mode, median, mean

Note: F, M = female, male; S, M, L = small, medium, large.

measured initially before we decide which statistical techniques or procedures to apply. Table 1.1 describes the levels of measurement and their associated measures of central tendency. Measures of central tendency are discussed in Chapter 3.

1.6.3 Sample Size

Sample size is an important criterion when choosing a particular statistical test or technique for data analysis. Most statistical tests have recommended sample sizes for their optimum application to collected data. Statistical tests are highly sensitive to sample size variations. The relation between sample size and its effect on statistical test results or statistical power will be addressed in Chapter 7 and subsequent chapters as different statistical tests and techniques are considered.

1.6.4 Other Assumptions

The assumptions discussed are fundamental to and relevant for most statistical test applications. Other assumptions may exist for certain statistical techniques, however. At these times, such assumptions will be explained and illustrated. For instance, distributional assumptions underlie the appropriate application of particular statistical tests. As we will see in Chapter 5, one assumption is that an investigator's sample scores should be normally distributed or distributed in a bell-shaped pattern. Many distributions of raw scores are not shaped in this fashion, and as a result, this distribution assumption is not met. Other statistical procedures require that if two or more samples are being compared, then all distributions of scores being compared must have a similar distribution pattern. When the different distributions of scores lack these distributional similarities, their application is inappropriate. These distributional assumptions will be clarified when the appropriate statistical tests involving them are presented and discussed.

1.7 Summary

- Several functions of statistical procedures include describing collected data, making decisions about hypothesis tests, and making inferences about population values based on sample characteristics.