Introduction to Biostatistics

1

CHAPTER OVERVIEW

This chapter is intended to provide an overview of the basic statistical concepts and definitions used throughout the textbook. A course in statistics requires the student to learn new and specific terminology. Therefore, this chapter lays the foundation necessary for understanding basic statistical terms and concepts and the role that statisticians play in promoting scientific discovery.

TOPICS

LEARNING OUTCOMES

After studying this chapter, the student will

1. understand the basic concepts and terminology of biostatistics, including types of variables, measurement, and measurement scales.
2. be able to select a simple random sample and other scientific samples from a population of subjects.
3. understand the processes involved in the scientific method.
4. appreciate the advantages of using computers in the statistical analysis of data generated by studies and experiments conducted by researchers in the health sciences.

1.1 Introduction

We are frequently reminded of the fact that we are living in the information age. Appropriately, then, this book is about information—how it is obtained, how it is analyzed, and how it is interpreted. The information about which we are concerned we call data, and the data are available to us in the form of numbers or in other non numerical forms that can be analyzed.

The objectives of this book are twofold: (1) to teach the student to organize and summarize data and (2) to teach the student how to reach decisions about a large body of data by examining only a small part of it. The concepts and methods necessary for achieving the first objective are presented under the heading of *descriptive statistics*, and the second objective is reached through the study of what is called *inferential statistics*. This chapter discusses descriptive statistics. Chapters 2 through 5 discuss topics that form the foundation of statistical inference, and most of the remainder of the book deals with inferential statistics.

Because this volume is designed for persons preparing for or already pursuing a career in the health field, the illustrative material and exercises reflect the problems and activities that these persons are likely to encounter in the performance of their duties.

1.2 Basic Concepts and Definitions

Like all fields of learning, statistics has its own vocabulary. Some of the words and phrases encountered in the study of statistics will be new to those not previously exposed to the subject. Other terms, though appearing to be familiar, may have specialized meanings that are different from the meanings that we are accustomed to associating with these terms. The following are some common terms that we will use extensively in this book; others will be added as we progress through the material.

Data

The raw material of statistics is *data*. For our purposes, we may define data as *numbers*. The two kinds of numbers that we use in statistics are numbers that result from the taking—in the usual sense of the term—of a *measurement*, and those that result from the process of *counting*. For example, when a nurse weighs a patient or takes a patient's temperature, a measurement, consisting of a number such as 150 pounds or 100 degrees Fahrenheit, is obtained. Quite a different type of number is obtained when a hospital administrator counts the number of patients—perhaps 20—discharged from the hospital on a given day. Each of the three numbers is a *datum*, and the three taken together are data. Data can also be understood to be non numerical, and may include things such as text or other qualitative items. However, we will focus our interests in this text largely on numerical data and their associated analyses.

Statistics

The meaning of *statistics* is implicit in the previous section. More concretely, however, we may say that *statistics is a field of study concerned with* (1) *the collection, organization, summarization, and analysis of data and* (2) *the drawing of inferences about a body of data when only a part of the data is observed*.

The person who performs these statistical activities must be prepared to *interpret* and to *communicate* the results to someone else as the situation demands. Simply put, we may say that data are numbers, numbers contain information, and the purpose of statistics is to investigate and evaluate the nature and meaning of this information.

Sources of Data

The performance of statistical activities is motivated by the need to answer a question. For example, clinicians may want answers to questions regarding the relative merits of competing treatment procedures. Administrators may want answers to questions regarding such areas of concern as employee morale or facility utilization. When we determine that the appropriate approach to seeking an answer to a question will require the use of statistics, we begin to search for suitable data to serve as the raw material for our investigation. Such data are usually available from one or more of the following sources:

1. **Routinely kept records.** It is difficult to imagine any type of organization that does not keep records of day-to-day transactions of its activities. Hospital medical records, for example, contain immense amounts of information on patients, while hospital accounting records contain a wealth of data on the facility's business activities. When the need for data arises, we should look for them first among routinely kept records.
2. **Surveys.** If the data needed to answer a question are not available from routinely kept records, the logical source may be a survey. Suppose, for example, that the administrator of a clinic wishes to obtain information regarding the mode of transportation used by patients to visit the clinic. If admission forms do not contain a question on mode of transportation, we may conduct a survey among patients to obtain this information.
3. **Experiments.** Frequently, the data needed to answer a question are available only as the result of an experiment. A nurse may wish to know which of several strategies is best for maximizing patient compliance. The nurse might conduct an experiment in which the different strategies of motivating compliance are tried with different patients. Subsequent evaluation of the responses to the different strategies might enable the nurse to decide which is most effective.
4. **External sources.** The data needed to answer a question may already exist in the form of published reports, commercially available data banks, or the research literature. In other words, we may find that someone else has already asked the same question, and the answer obtained may be applicable to our present situation.

Biostatistics

The tools of statistics are employed in many fields—business, education, psychology, agriculture, and economics, to mention only a few. When the data analyzed are derived from the biological sciences and medicine, we use the term *biostatistics* to distinguish this particular application of statistical tools and concepts. This area of application is the concern of this book.

Variable

If, as we observe a characteristic, we find that it takes on different values in different persons, places, or things, we label the characteristic a *variable*. We do this for the simple reason that the characteristic is not the same when observed in different possessors of it. Some examples of variables include diastolic blood pressure, heart rate, the heights of adult males, the weights of preschool children, and the ages of patients seen in a dental clinic.

Quantitative Variables

A *quantitative variable* is one that can be measured in the usual sense. We can, for example, obtain measurements on the heights of adult males, the weights of preschool children, and the ages of

patients seen in a dental clinic. These are examples of *quantitative variables*. Measurements made on quantitative variables convey information regarding amount.

Qualitative Variables

Some characteristics are not capable of being measured in the sense that height, weight, and age are measured. Many characteristics can be categorized only, as, for example, when an ill person is given a medical diagnosis, a person is designated as belonging to an ethnic group, or a person, place, or object is said to possess or not to possess some characteristic of interest. In such cases, measuring consists of categorizing. We refer to variables of this kind as *qualitative variables*. Measurements made on qualitative variables convey information regarding an attribute.

Although, in the case of qualitative variables, measurement in the usual sense of the word is not achieved, we can count the number of persons, places, or things belonging to various categories. A hospital administrator, for example, can count the number of patients admitted during a day under each of the various admitting diagnoses. These counts, or *frequencies* as they are called, are the numbers that we manipulate when our analysis involves qualitative variables.

Random Variable

Whenever we determine the height, weight, or age of an individual, the result is frequently referred to as a *value* of the respective variable. When the values obtained arise as a result of chance factors, so that they cannot be exactly predicted in advance, the variable is called a *random variable*. An example of a random variable is adult height. When a child is born, we cannot predict exactly his or her height at maturity. Attained adult height is the result of numerous genetic and environmental factors. Values resulting from measurement procedures are often referred to as *observations* or *measurements*.

Discrete Random Variable

Variables may be characterized further as to whether they are *discrete* or *continuous*. Since mathematically rigorous definitions of discrete and continuous variables are beyond the level of this book, we offer, instead, nonrigorous definitions and give an example of each.

A discrete variable is characterized by gaps or interruptions in the values that it can assume. These gaps or interruptions indicate the absence of values between particular values that the variable can assume. Some examples illustrate the point. The number of daily admissions to a general hospital is a discrete random variable since the number of admissions each day must be represented by a whole number, such as 0, 1, 2, or 3. The number of admissions on a given day cannot be a number such as 1.5, 2.997, or 3.333. The number of decayed, missing, or filled teeth per child in an elementary school is another example of a discrete variable.

Continuous Random Variable

A continuous random variable does not possess the gaps or interruptions characteristic of a discrete random variable. A continuous random variable can assume any value within a specified relevant interval of values assumed by the variable. Examples of continuous variables include the various measurements that can be made on individuals such as height, weight, and skull circumference. No matter how close together the observed heights of two people, for example, we can, theoretically, find another person whose height falls somewhere in between.

Because of the limitations of available measuring instruments, however, observations on variables that are inherently continuous are recorded as if they were discrete. Height, for example, is usually recorded to the nearest one-quarter, one-half, or whole inch, whereas, with a perfect measuring device, such a measurement could be made as precise as desired. Therefore, in a nontechnical sense, continuity is limited only by our ability to precisely measure it.

Population

The average person thinks of a population as a collection of entities, usually people. A population or collection of entities may, however, consist of animals, machines, places, or cells. For our purposes, we define a *population of entities as the largest collection of entities for which we have an interest at a particular time*. If we take a measurement of some variable on each of the entities in a population, we generate a population of values of that variable. We may, therefore, define a *population of values as the largest collection of values of a random variable for which we have an interest at a particular time*. If, for example, we are interested in the weights of all the children enrolled in a certain county elementary school system, our population consists of all these weights. If our interest lies only in the weights of first-grade students in the system, we have a different population—weights of first-grade students enrolled in the school system. Hence, populations are determined or defined by our sphere of interest. Populations may be *finite* or *infinite*. If a population of values consists of a fixed number of these values, the population is said to be *finite*. If, on the other hand, a population consists of an endless succession of values, the population is an *infinite* one. An exact value calculated from a population is referred to as a *parameter*.

Sample

A sample may be defined simply as *a part of a population*. Suppose our population consists of the weights of all the elementary school children enrolled in a certain county school system. If we collect for analysis the weights of only a fraction of these children, we have only a part of our population of weights, that is, we have a *sample*. An estimated value calculated from a sample is referred to as a *statistic*.

1.3 Measurement and Measurement Scales

In the preceding discussion, we used the word *measurement* several times in its usual sense, and presumably the reader clearly understood the intended meaning. The word *measurement*, however, may be given a more scientific definition. In fact, there is a whole body of scientific literature devoted to the subject of measurement. Part of this literature is concerned also with the nature of the numbers that result from measurements. Authorities on the subject of measurement speak of measurement scales that result in the categorization of measurements according to their nature. In this section, we define measurement and the four resulting measurement scales. A more detailed discussion of the subject is to be found in the writings of Stevens (1,2).

Measurement

This may be defined as the assignment of numbers to objects or events according to a set of rules. The various measurement scales result from the fact that measurement may be carried out under different sets of rules.

The Nominal Scale

The lowest measurement scale is the *nominal scale*. As the name implies it consists of "naming" observations or classifying them into various mutually exclusive and collectively exhaustive categories. The practice of using numbers to distinguish among the various medical diagnoses constitutes measurement on a nominal scale. Other examples include such dichotomies as positive–negative, well–sick, under 65 years of age–65 and over, child–adult, and married–not married.

The Ordinal Scale

Whenever observations are not only different from category to category but can be ranked according to some criterion, they are said to be measured on an ordinal scale. Convalescing patients may be characterized as unimproved, improved, and much improved. Individuals may be classified according to socioeconomic status as low, medium, or high. The intelligence of children may be above average, average, or below average. In each of these examples, the members of any one category are all considered equal, but the members of one category are considered lower, worse, or smaller than those in another category, which in turn bears a similar relationship to another category. For example, a much improved patient is in better health than one classified as improved, while a patient who has improved is in better condition than one who has not improved. It is usually impossible to infer that the difference between members of one category and the next adjacent category is equal to the difference between members of that category and the members of the next category adjacent to it. The degree of improvement between unimproved and improved is probably not the same as that between improved and much improved. The implication is that if a finer breakdown were made resulting in more categories, these, too, could be ordered in a similar manner. The function of numbers assigned to ordinal data is to order (or rank) the observations from lowest to highest and, hence, the term *ordinal*.

The Interval Scale

The *interval scale* is a more sophisticated scale than the nominal or ordinal in that with this scale not only is it possible to order measurements, but also the distance between any two measurements is known. We know, say, that the difference between a measurement of 20 and a measurement of 30 is equal to the difference between measurements of 30 and 40. The ability to do this implies the use of a unit distance and a zero point, both of which are arbitrary. The selected zero point is not necessarily a true zero in that it does not have to indicate a total absence of the quantity being measured. Perhaps the best example of an interval scale is provided by the way in which temperature is usually measured (degrees Fahrenheit or Celsius). The unit of measurement is the degree, and the point of comparison is the arbitrarily chosen "zero degrees," which does not indicate a lack of heat. The interval scale unlike the nominal and ordinal scales is a truly quantitative scale.

The Ratio Scale

The highest level of measurement is the *ratio scale*. This scale is characterized by the fact that equality of ratios as well as equality of intervals may be determined. Fundamental to the ratio scale is a true zero point. The measurement of such familiar traits as height, weight, and length makes use of the ratio scale.

1.4 Sampling and Statistical Inference

As noted earlier, one of the purposes of this book is to teach the concepts of statistical inference, which we may define as follows:

DEFINITION

Statistical inference is the procedure by which we reach a conclusion about a population on the basis of the information contained in a sample that has been drawn from that population.

There are many kinds of samples that may be drawn from a population. Not every kind of sample, however, can be used as a basis for making valid inferences about a population. In general, in order to make a valid inference about a population, we need a scientific sample from the population. There are also many kinds of scientific samples that may be drawn from a population. The simplest of these is the *simple random sample*. In this section, we define a simple random sample and show you how to draw one from a population.

If we use the letter N to designate the size of a finite population and the letter n to designate the size of a sample, we may define a simple random sample as follows:

DEFINITION

If a sample of size n is drawn from a population of size N in such a way that every possible sample of size n has the same chance of being selected, the sample is called a simple random sample.

The mechanics of drawing a sample to satisfy the definition of a simple random sample is called *simple random sampling*.

We will demonstrate the procedure of simple random sampling shortly, but first let us consider the problem of whether to sample *with replacement* or *without replacement*. When sampling with replacement is employed, every member of the population is available at each draw. For example, suppose that we are drawing a sample from a population of former hospital patients as part of a study of length of stay. Let us assume that the sampling involves selecting from the electronic health records a sample of charts of discharged patients. In sampling with replacement we would proceed as follows: select a chart to be in the sample, record the length of stay, and close the electronic chart. The chart is back in the "population" and may be selected again on some subsequent draw, in which case the length of stay will again be recorded. In sampling without replacement, we would not record a length of stay for a patient whose chart was already selected from the database. Following this procedure, a given chart could appear in the sample only once. In practice, sampling is almost always done without replacement. The significance and consequences of this will be explained later, but first let us see how one goes about selecting a simple random sample. To ensure true randomness of selection, we will need to follow some objective procedure. We certainly will want to avoid using our own judgment to decide which members of the population constitute a random sample. The following example illustrates one method of selecting a simple random sample from a population.

EXAMPLE 1.4.1

Gold et al. (A-1) studied the effectiveness on smoking cessation of bupropion SR, a nicotine patch, or both, when co administered with cognitive behavioral therapy. Consecutive consenting patients assigned themselves to one of the three conditions. For illustrative purposes, let us

consider all these subjects to be a population of size $N = 189$. We wish to select a simple random sample of size 10 from this population whose ages are shown in Table 1.4.1.

Table 1.4.1 Ages of 189 Subjects Who Participated in a Study on Smoking Cessation

Subject No.	Age	Subject No.	Age	Subject No.	Age	Subject No.	Age
1	48	49	38	97	51	145	52
2	35	50	44	98	50	146	53
3	46	51	43	99	50	147	61
4	44	52	47	100	55	148	60
5	43	53	46	101	63	149	53
6	42	54	57	102	50	150	53
7	39	55	52	103	59	151	50
8	44	56	54	104	54	152	53
9	49	57	56	105	60	153	54
10	49	58	53	106	50	154	61
11	44	59	64	107	56	155	61
12	39	60	53	108	68	156	61
13	38	61	58	109	66	157	64
14	49	62	54	110	71	158	53
15	49	63	59	111	82	159	53
16	53	64	56	112	68	160	54
17	56	65	62	113	78	161	61
18	57	66	50	114	66	162	60
19	51	67	64	115	70	163	51
20	61	68	53	116	66	164	50
21	53	69	61	117	78	165	53
22	66	70	53	118	69	166	64
23	71	71	62	119	71	167	64
24	75	72	57	120	69	168	53
25	72	73	52	121	78	169	60
26	65	74	54	122	66	170	54
27	67	75	61	123	68	171	55
28	38	76	59	124	71	172	58
29	37	77	57	125	69	173	62
30	46	78	52	126	77	174	62
31	44	79	54	127	76	175	54
32	44	80	53	128	71	176	53
33	48	81	62	129	43	177	61
34	49	82	52	130	47	178	54
35	30	83	62	131	48	179	51
36	45	84	57	132	37	180	62

Table 1.4.1 (Continued)

Subject No.	Age	Subject No.	Age	Subject No.	Age	Subject No.	Age
37	47	85	59	133	40	181	57
38	45	86	59	134	42	182	50
39	48	87	56	135	38	183	64
40	47	88	57	136	49	184	63
41	47	89	53	137	43	185	65
42	44	90	59	138	46	186	71
43	48	91	61	139	34	187	71
44	43	92	55	140	46	188	73
45	45	93	61	141	46	189	66
46	40	94	56	142	48		
47	48	95	52	143	47		
48	49	96	54	144	43		

Source: Data provided courtesy of Paul B. Gold, Ph.D.

SOLUTION: One way of selecting a simple random sample is to use a table of random numbers, which was very commonly done historically, or to use an online random number generator. However, most statistical computer packages provide a way to select a sample of given size. For example, using program R, we can create a sequence of numbers from 1 to 189, and call them "subject" and then use the sample() function to randomly select 10 of them. One such output is shown in Figure 1.4.1, with the output summarized in Table 1.4.2.

```
> subject <- seq(1,189,1)
> sample(subject,10)
 [1] 151 106  61 119  44  88  75  68  20  94
```

FIGURE 1.4.1 Simple R program for selecting a sample of size 10 from 189 subjects.

Table 1.4.2 Sample of 10 Ages Drawn from the Ages in Table 1.4.1

Random Number	Sample Subject Number	Age
151	1	50
106	2	50
61	3	58
119	4	71
44	5	79
88	6	57
75	7	61
68	8	53
20	9	61
94	10	56

Thus we have drawn a simple random sample of size 10 from a population of size 189. In future discussions, whenever the term *simple random sample* is used, it will be understood that the sample has been drawn in this or an equivalent manner.

The preceding discussion of random sampling is presented because of the important role that the sampling process plays in designing *research studies* and *experiments*. The methodology and concepts employed in sampling processes will be described in more detail in Section 1.5.

DEFINITION

A research study is a scientific study of a phenomenon of interest. Research studies involve designing sampling protocols, collecting and analyzing data, and providing valid conclusions based on the results of the analyses.

DEFINITION

Experiments are a special type of research study in which observations are made after specific manipulations of conditions have been carried out; they provide the foundation for scientific research.

Despite the tremendous importance of random sampling in the design of research studies and experiments, there are some occasions when random sampling may not be the most appropriate method to use. Consequently, other sampling methods must be considered. The intention here is not to provide a comprehensive review of sampling methods, but rather to acquaint the student with two additional sampling methods that are often employed in the health sciences, *systematic sampling* and *stratified random sampling*. Interested readers are referred to the books by Thompson (3) and Levy and Lemeshow (4) for detailed overviews of various sampling methods and explanations of how sample statistics are calculated when these methods are applied in research studies and experiments.

Systematic Sampling

A sampling method that is widely used in health-care research is the systematic sample. Medical records, which contain raw data used in health-care research, are generally stored in a file system or on a computer and hence are easy to select in a systematic way. Using systematic sampling methodology, a researcher calculates the total number of records needed for the study or experiment at hand. A random number is then chosen to use as a starting point for initiating sampling. The record located at this starting point is called record x. A second number, determined by the number of records desired, is selected to define the sampling interval (call this interval k). Consequently, the data set would consist of records $x, x + k, x + 2k, x + 3k$, and so on, until the necessary number of records are obtained.

EXAMPLE 1.4.2

Continuing with the study of Gold et al. (A-1) illustrated in the previous example, imagine that we wanted a systematic sample of 10 subjects from those listed in Table 1.4.1.

SOLUTION: To obtain a starting point, we can employ the strategy from above using R and simply select only one subject. Let us assume that the sample() function returned subject number 4, which will serve as our starting point, x. Since we are starting at subject 4, this leaves 185 remaining subjects (i.e., 189 − 4) from which to choose. Since we wish to select 10 subjects, one method to define the sample interval, k, would be to take 185/10 = 18.5. To ensure that there will be enough subjects, it is customary to round this quotient down, and hence we will round the result to 18. In this scenario, the samples would be 4, 4 + 18 = 22, 4 + 18(2) = 40, 4 + 18(3) = 58, and so on. The resulting sample is shown in Table 1.4.3.

Table 1.4.3 Sample of 10 Ages Selected Using a Systematic Sample from the Ages in Table 1.4.1

Systematically Selected Subject Number	Age
4	44
22	66
40	47
58	53
76	59
94	56
112	68
130	47
148	60
166	64

Stratified Random Sampling

A common situation that may be encountered in a population under study is one in which the sample units occur together in a grouped fashion. On occasion, when the sample units are not inherently grouped, it may be possible and desirable to group them for sampling purposes. In other words, it may be desirable to partition a population of interest into groups, or *strata*, in which the sample units within a particular stratum are more similar to each other than they are to the sample units that compose the other strata. After the population is stratified, it is customary to take a random sample independently from each stratum. This technique is called *stratified random sampling*. The resulting sample is called a *stratified random sample*. Although the benefits of stratified random sampling may not be readily observable, it is most often the case that random samples taken within a stratum will have much less variability than a random sample taken across all strata. This is true because sample units within each stratum tend to have characteristics that are similar.

EXAMPLE 1.4.3

Hospital trauma centers are given ratings depending on their capabilities to treat various traumas. In this system, a level 1 trauma center is the highest level of available trauma care and a level 4 trauma center is the lowest level of available trauma care. Imagine that we are interested in estimating the survival rate of trauma victims treated at hospitals within a large metropolitan area. Suppose that the metropolitan area has a level 1, a level 2, and a level 3 trauma center. We

wish to take samples of patients from these trauma centers in such a way that the total sample size is 30.

SOLUTION: We assume that the survival rates of patients may depend quite significantly on the trauma that they experienced and therefore on the level of care that they receive. As a result, a simple random sample of all trauma patients, without regard to the center at which they were treated, may not represent true survival rates, since patients receive different care at the various trauma centers. One way to better estimate the survival rate is to treat each trauma center as a stratum and then randomly select 10 patient files from each of the three centers. This procedure is based on the fact that we suspect that the survival rates within the trauma centers are less variable than the survival rates across trauma centers. Therefore, we believe that the stratified random sample provides a better representation of survival than would a sample taken without regard to differences within strata.

It should be noted that two slight modifications of the stratified sampling technique are frequently employed. To illustrate, consider again the trauma center example. In the first place, a systematic sample of patient files could have been selected from each trauma center (stratum). Such a sample is called a *stratified systematic sample*.

The second modification of stratified sampling involves selecting the sample from a given stratum in such a way that the number of sample units selected from that stratum is proportional to the size of the population of that stratum. Suppose, in our trauma center example that the level 1 trauma center treated 100 patients and the level 2 and level 3 trauma centers treated only 10 each. In that case, selecting a random sample of 10 from each trauma center overrepresents the trauma centers with smaller patient loads. To avoid this problem, we adjust the size of the sample taken from a stratum so that it is proportional to the size of the stratum's population. This type of sampling is called *stratified sampling proportional to size*. The within-stratum samples can be either random or systematic as described above.

Exercises

1.4.1 Using a table of random numbers or a computer program, select a new simple random sample of size 10 from the data in Table 1.4.1. Record the ages of the subjects in this new sample. Save your data for future use. What is the variable of interest in this exercise? What measurement scale was used to obtain the measurements?

1.4.2 Select another simple random sample of size 10 from the population represented in Table 1.4.1. Compare the subjects in this sample with those in the sample drawn in Exercise 1.4.1. Are there any subjects who showed up in both samples? How many? Compare the ages of the subjects in the two samples. How many ages in the first sample were duplicated in the second sample?

1.4.3 Using a table of random numbers or a computer program, select a random sample and a systematic sample, each of size 15, from the data in Table 1.4.1. Visually compare the distributions of the two samples. Do they appear similar? Which appears to be the best representation of the data?

1.4.4 Construct an example where it would be appropriate to use stratified sampling. Discuss how you would use stratified random sampling and stratified sampling proportional to size with this example. Which do you think would best represent the population that you described in your example? Why?

1.5 The Scientific Method

Data analyses using a broad range of statistical methods play a significant role in scientific studies. The previous section highlighted the importance of obtaining samples in a scientific manner. Appropriate sampling techniques enhance the likelihood that the results of statistical analyses of a data set will provide valid and scientifically defensible results. Because of the importance of the proper collection of data to support scientific discovery, it is necessary to consider the foundation of such discovery—the *scientific method*—and to explore the role of statistics in the context of this method.

DEFINITION

The scientific method is a process by which scientific information is collected, analyzed, and reported in order to produce unbiased and replicable results in an effort to provide an accurate representation of observable phenomena.

The scientific method is recognized universally as the only truly acceptable way to produce new scientific understanding of the world around us. It is based on an *empirical approach*, in that decisions and outcomes are based on data. There are several key elements associated with the scientific method, and the concepts and techniques of statistics play a prominent role in all these elements. The Scientific Method is illustrated in Figure 1.5.1.

Observations and Question Formation

First, an *observation* is made of a phenomenon or a group of phenomena. This observation leads to the formulation of questions or uncertainties that can be answered in a scientifically rigorous way. For example, it is readily observable that regular exercise reduces body weight in many people. It is also readily observable that changing diet may have a similar effect. In this case, there are two observable phenomena, regular exercise and diet change, that have the same endpoint. The nature of this endpoint can be determined by use of the scientific method.

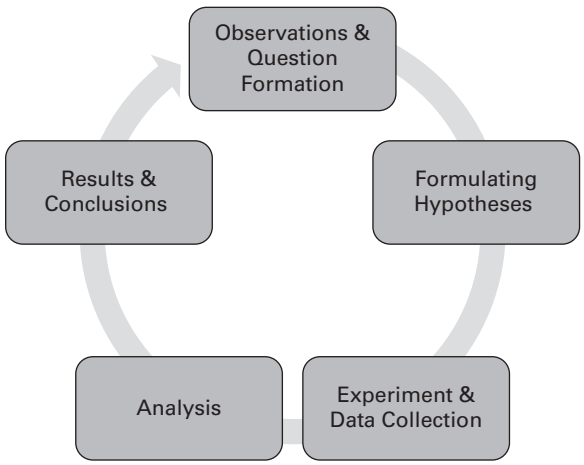

FIGURE 1.5.1 An illustration of the Scientific Method.

Formulating Hypotheses

In the second step of the scientific method, a *hypothesis* is formulated to explain the observation and to make quantitative *predictions* of new observations. Often hypotheses are generated as a result of extensive background research and literature reviews. The objective is to produce hypotheses that are scientifically sound. Hypotheses may be stated as either *research hypotheses* or *statistical hypotheses*. Explicit definitions of these terms are given in Chapter 7, which discusses the science of testing hypotheses. Suffice it to say for now that a research hypothesis from the weight-loss example would be a statement such as "Exercise appears to reduce body weight." There is certainly nothing incorrect about this conjecture, but it lacks a truly quantitative basis for testing. A statistical hypothesis may be stated using quantitative terminology as follows: "The average (mean) loss of body weight of people who exercise is greater than the average (mean) loss of body weight of people who do not exercise." In this statement a quantitative measure, the "average" or "mean" value, is hypothesized to be greater in the sample of patients who exercise. The role of the statistician in this step of the scientific method is to state the hypothesis in a way that valid conclusions may be drawn and to interpret correctly the results of such conclusions.

Experiment and Data Collection

The third step of the scientific method involves *designing an experiment* that will yield the data necessary to validly test an appropriate statistical hypothesis. This step of the scientific method, like that of data analysis, requires the expertise of a statistician. Improperly designed experiments are the leading cause of invalid results and unjustified conclusions. Further, most studies that are challenged by experts are challenged on the basis of the appropriateness or inappropriateness of the study's research design, which can lead to a nonreproducible result.

Those who properly design research experiments make every effort to ensure that the measurement of the phenomenon of interest is both accurate and precise. *Accuracy* refers to the correctness of a measurement. *Precision*, on the other hand, refers to the consistency of a measurement. It should be noted that in the social sciences, the term *validity* is sometimes used to mean accuracy and that *reliability* is sometimes used to mean precision. In the context of the weight-loss example given earlier, the scale used to measure the weight of study participants would be accurate if the measurement is validated using a scale that is properly calibrated. If, however, the scale is off by +3 pounds, then each participant's weight would be 3 pounds heavier; the measurements would be precise in that each would be wrong by +3 pounds, but the measurements would not be accurate. Measurements that are inaccurate or imprecise may invalidate research findings.

The design of an experiment depends on the type of data that need to be collected to test a specific hypothesis. As discussed in Section 1.2, data may be collected or made available through a variety of means. For much scientific research, however, the standard for data collection is experimentation. A true *experimental design* is one in which study subjects are randomly assigned to an *experimental group* (or *treatment group*) and a *control group* that is not directly exposed to a treatment. Continuing the weight-loss example, a sample of 100 participants could be randomly assigned to two conditions using the methods of Section 1.4. A sample of 50 of the participants would be assigned to a specific exercise program and the remaining 50 would be monitored, but asked not to exercise for a specific period of time. At the end of this experiment, the average (mean) weight losses of the two groups could be compared. The reason that experimental designs are desirable is that if all other potential factors are controlled, a *cause–effect relationship* may be tested; that is, all else being equal, we would be able to conclude or fail to conclude that the experimental group lost weight as a result of exercising.

The potential complexity of research designs requires statistical expertise, and Chapter 8 highlights some commonly used experimental designs. For a more in-depth discussion of research

designs, the interested reader may wish to refer to texts by Kuehl (5), Keppel and Wickens (6), and Tabachnick and Fidell (7).

Analysis

Assuming that an appropriate experimental design is employed and data are correctly collected through a validated sampling protocol, these resulting data can be analyzed. Often *descriptive statistics*, as we will learn in Chapter 2, are used to understand characteristics of the data that have been collected. Additionally, based on the hypotheses that were developed, a researcher may have one or more analyses to complete using *inferential statistics*, as covered starting in Chapter 7, in order to make predictions or infer conclusions from the data.

Results and Conclusions

In the execution of a research study or experiment, one would hope to have collected the data necessary to draw conclusions, with some degree of confidence, about the hypotheses that were posed as part of the design. It is often the case that hypotheses need to be modified and retested with new data and a different design. Whatever the conclusions of the scientific process, however, results are rarely considered to be conclusive. That is, results need to be *replicated*, often a large number of times, before scientific credence is granted them.

Exercises

1.5.1 Using the example of weight loss as an endpoint, discuss how you would use the scientific method to test the observation that change in diet is related to weight loss. Include all of the steps, including the hypothesis to be tested and the design of your experiment.

1.5.2 Continuing with Exercise 1.5.1, consider how you would use the scientific method to test the observation that both exercise and change in diet are related to weight loss. Include all of the steps, paying particular attention to how you might design the experiment and which hypotheses would be testable given your design.

1.6 Computers and Technology

The widespread use of computers and related technology has had a tremendous impact on health sciences research in general and biostatistical analysis in particular. The necessity to perform long and tedious arithmetic computations as part of the statistical analysis of data lives only in the memory of those researchers and practitioners whose careers antedate the so-called computer revolution. Likewise, the use of statistical tables for finding standard comparative (i.e., critical) values in hypothesis testing largely has been replaced with statistical algorithms. Computers can perform more calculations faster and far more accurately than can human technicians. The use of computers makes it possible for investigators to devote more time to the improvement of the quality of raw data and the interpretation of the results.

The current prevalence of microcomputers and the abundance of available statistical software programs have further revolutionized statistical computing. The reader in search of a statistical software package may wish to consult *The American Statistician*, a quarterly publication of

the American Statistical Association. Statistical software packages are regularly reviewed and advertised in the periodical.

As was illustrated in a previous example, an alternative to using printed tables of random numbers, investigators may use computers to generate the random numbers they need. Actually, the "random" numbers generated by most computers are in reality *pseudorandom numbers* because they are the result of a deterministic formula. However, as Fishman (8) points out, the numbers appear to serve satisfactorily for many practical purposes.

The usefulness of the computer in the health sciences is not limited to statistical analysis. The reader interested in learning more about the use of computers in the health sciences will find the books by Hersh (9), Johns (10), Miller et al. (11), and Saba and McCormick (12) helpful; though some of these are now dated, the perspectives discussed in these volumes is still relevant. Those who wish to derive maximum benefit from the Internet may wish to consult the books *Physicians' Guide to the Internet* (13) and *Computers in Nursing's Nurses' Guide to the Internet* (14). Current developments in the use of computers in biology, medicine, and related fields are reported in several periodicals devoted to the subject. A few such periodicals are *Computers in Biology and Medicine*, *Computers and Biomedical Research*, *International Journal of Bio-Medical Computing*, *Computer Methods and Programs in Biomedicine*, *Computer Applications in the Biosciences*, and *Computers in Nursing*.

As the reader makes their way through this text, it will be apparent just how important computers and technology have become to our current use of statistics for applied problems. We will generally forgo lengthy hand calculations and tabled values in favor of a more contemporary view of statistical science. In that regard, Computer printouts are used throughout this book to illustrate the use of computers in biostatistical analysis. The MINITAB, SPSS, R, JASP, and SAS® statistical software packages for the personal computer have been used for this purpose. Additionally, we will sometimes present some useful Microsoft Excel™ outputs as well.

1.7 SUMMARY

In this chapter, we introduced the reader to the basic concepts of statistics. We defined statistics as an area of study concerned with collecting and describing data and with making statistical inferences. We defined statistical inference as the procedure by which we reach a conclusion about a population on the basis of information contained in a sample drawn from that population. We learned that a basic type of sample that will allow us to make valid inferences is the simple random sample. We learned how to use a table of random numbers to draw a simple random sample from a population.

The reader is provided with the definitions of some basic terms, such as variable and sample, that are used in the study of statistics. We also discussed measurement and defined four measurement scales—nominal, ordinal, interval, and ratio. The reader is also introduced to the scientific method and the role of statistics and the statistician in this process.

Finally, we discussed the importance of computers in the performance of the activities involved in statistics.

REVIEW QUESTIONS AND EXERCISES

1. Explain what is meant by descriptive statistics.

2. Explain what is meant by inferential statistics.

3. Define:

- **(a)** Statistics
- **(b)** Biostatistics
- **(c)** Variable
- **(d)** Quantitative variable
- **(e)** Qualitative variable
- **(f)** Random variable
- **(g)** Population
- **(h)** Finite population
- **(i)** Infinite population

(j) Sample

(k) Discrete variable

(l) Continuous variable

(m) Simple random sample

(n) Sampling with replacement

(o) Sampling without replacement

4. Define the word *measurement*.

5. List, describe, and compare the four measurement scales.

6. For each of the following variables, indicate whether it is quantitative or qualitative and specify the measurement scale that is employed when taking measurements on each:

(a) Class standing of the members of this class relative to each other

(b) Admitting diagnosis of patients admitted to a mental health clinic

(c) Weights of babies born in a hospital during a year

(d) Gender of babies born in a hospital during a year

(e) Range of motion of elbow joint of students enrolled in a university health sciences curriculum

(f) Under-arm temperature of day-old infants born in a hospital

7. For each of the following situations, answer questions a through e:

(a) What is the sample in the study?

(b) What is the population?

(c) What is the variable of interest?

(d) How many measurements were used in calculating the reported results?

(e) What measurement scale was used?

Situation A. A study of 300 households in a small southern town revealed that 20 percent had at least one school-age child present.

Situation B. A study of 250 patients admitted to a hospital during the past year revealed that, on the average, the patients lived 15 miles from the hospital.

8. Consider the two situations given in Exercise 7. For Situation A, describe how you would use a stratified random sample to collect the data. For Situation B, describe how you would use systematic sampling of patient records to collect the data.

REFERENCES

Methodology References

1. Stanley S. Stevens, "On the Theory of Scales of Measurement," *Science*, 103 (1946), 677–680.
2. Stanley S. Stevens, "Mathematics, Measurement and Psychophysics," in Stanley S. Stevens (ed.), *Handbook of Experimental Psychology*, Wiley, New York, 1951.
3. Steven K. Thompson, *Sampling* (3rd ed.), Wiley, New York, 2012.
4. Paul S. Levy and Stanley Lemeshow, *Sampling of Populations: Methods and Applications* (4th ed.), Wiley, New York, 2008.
5. Robert O. Kuehl, *Statistical Principles of Research Design and Analysis* (2nd ed.), Duxbury Press, Belmont, CA, 1999.
6. Geoffrey Keppel and Thomas D. Wickens, *Design and Analysis: A Researcher's Handbook* (4th ed.), Prentice Hall, Upper Saddle River, NJ, 2004.
7. Barbara G. Tabachnick and Linda S. Fidell, *Experimental Designs using ANOVA*, Thomson, Belmont, CA, 2007.
8. George S. Fishman, *Concepts and Methods in Discrete Event Digital Simulation*, Wiley, New York, 1973.
9. William R. Hersh, *Information Retrieval: A Health Care Perspective* (3rd ed.), Springer, New York, 2010.
10. Merida L. Johns, *Information Management for Health Professions*, Delmar Publishers, Albany, NY, 1997.
11. Marvin J. Miller, Kenric W. Hammond, and Matthew G. Hile (eds.), *Mental Health Computing*, Springer, New York, 1996.
12. Virginia K. Saba and Kathleen A. McCormick, *Essentials of Computers for Nurses*, McGraw-Hill, New York, 1996.
13. Lee Hancock, *Physicians' Guide to the Internet*, Lippincott Williams & Wilkins Publishers, Philadelphia, 1996.
14. Leslie H. Nicoll and Teena H. Ouellette, *Computers in Nursing's Nurses' Guide to the Internet* (3rd ed.), Lippincott Williams & Wilkins Publishers, Philadelphia, 2001.

Applications References

A-1. Paul B. Gold, Robert N. Rubey, and Richard T. Harvey, "Naturalistic, Self-Assignment Comparative Trial of Bupropion SR, a Nicotine Patch, or Both for Smoking Cessation Treatment in Primary Care," *American Journal on Addictions*, 11 (2002), 315–331.