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 shelves in the medical records department 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 return the chart to the shelf. The chart is back in the “population” and may be drawn again on some subsequent draw, in which case the length of stay will again be recorded. In sampling without replacement, we would not return a drawn chart to the shelf after recording the length of stay, but would lay it aside until the entire sample is drawn. Following this procedure, a given chart could appear in the sample only once. As a rule, in practice, sampling is 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 treatments. 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.

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 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 healthcare research is the systematic sample. Medical records, which contain raw data used in healthcare 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 numbers table is then employed to select a starting point in the file system. 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, imag-

ine that we wanted a systematic sample of 10 subjects from those listed in Table 1.4.1.

**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.

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.

**1.5 THE SCIENTIFIC METHOD AND**

**THE DESIGN OF EXPERIMENTS**

Data analyses by 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.

**Making an Observation** 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.

**Formulating a Hypothesis** 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.

**Designing an Experiment** 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.

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 heavy; 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).

**Conclusion** 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.