Research Methodology / Master 1 - Didactics
Lecture 1: Sampling
Source: Introduction to Research in Education (Ary, et al., 2009: pp. 147-162)

SAMPLING

An important characteristic of inferential statistics is the process of going from the part to the whole. For example, you might study a randomly selected group of 500 students attending a university in order to make generalizations about the entire student body of that university.

The small group that is observed is called a *sample*, and the larger group about which the generalization is made is called a *population*. A **population** is defined as all members of any well-defined class of people, events, or objects. For example, in a study in which students in American high schools constitute the population of interest, you could define this population as all boys and girls attending high school in the United States. A **sample** is a portion of a population. For example, the students of Washington High School in Indianapolis constitute a sample of American high school students.

Statistical inference is a procedure by means of which you estimate **parameters** (characteristics of populations) from **statistics** (characteristics of samples). Such estimations are based on the laws of probability and are best estimates rather than absolute facts. In making any such inferences, a certain degree of error is involved. Inferential statistics can be used to test hypotheses about populations on the basis of observations of a sample drawn from the population.

RATIONALE OF SAMPLING

Inductive reasoning is an essential part of the scientific approach. The inductive method involves making observations and then drawing conclusions from these observations. If you can observe all instances of a population, you can, with confidence, base conclusions about the population on these observations (perfect induction). In Chapter 6, we treated the 18 students in Mr. Li's physics class as a population. Therefore, we could be confident that we had the true means, standard deviations, and so forth (the parameters). However, if you observe only some instances of a population, then you can do no more than infer that these observations will be true of the population as a whole (imperfect induction). This is the concept of sampling, which involves taking a portion of the population, making observations on this smaller group, and then generalizing the findings to the parent population—the larger population from which the sample was drawn.

Sampling is indispensable to the researcher. Usually, the time, money, and effort involved do not permit a researcher to study all possible members of a population. Furthermore, it is generally not necessary to study all possible cases to understand the phenomenon under consideration. Sampling comes to your aid by enabling you to study a portion of the population rather than the entire population.

Because the purpose of drawing a sample from a population is to obtain information concerning that population, it is extremely important that the individuals included in a sample constitute a representative cross section of individuals in the population. Samples must be representative if you are to be able to generalize with reasonable confidence from the sample to the population. For example, the researcher may assume that the students at Washington High School are representative of American adolescents. However, this sample may not be representative if the individuals who are included have some characteristics that differ from the target population. The location of their school, their socioeconomic backgrounds, their family situations, their prior experiences, and many other characteristics of this group may make them unrepresentative of American adolescents. An unrepresentative sample is termed a **biased sample.** The findings on a biased sample in a research study cannot legitimately be generalized to the population from which it is taken. For example, if the population of interest is all students in a particular urban school district but the researchers sampled only students from the district's two magnet schools, the sample would be biased.

STEPS IN SAMPLING

The first step in sampling is the identification of the **target population,** the large group to which the researcher wishes to generalize the results of the study. If the researcher is interested in learning about the teachers in the St. Louis public school system, all those who teach within that system constitute the target population. In a study of the attitudes and values of American adolescents, the target population would be all American boys and girls in the age range of 12 to 21 years, given that adolescence is operationally defined as the period between ages 12 and 21 years. We make a distinction between the target population and the **accessible population,** which is the population of subjects accessible to the researcher for drawing a sample. In most research, we deal with accessible populations. It would be expensive and time-consuming to sample from the total population of American adolescents, but we could draw a sample of adolescents from one state. Of course, we could only generalize results to adolescents in the chosen state, not to all American adolescents.

Once we have identified the population, the next step is to select the sample. Two major types of sampling procedures are available to researchers: probability and nonprobability sampling. **Probability sampling** involves sample selection in which the elements are drawn by chance procedures. The main characteristic of probability sampling is that every member or element of the population has a known probability of being chosen in the sample.

Nonprobability sampling includes methods of selection in which elements are not chosen by chance procedures. Its success depends on the knowledge, expertise, and judgment of the researcher. Nonprobability sampling is used when the application of probability sampling is not feasible. Its advantages are convenience and economy.

PROBABILITY SAMPLING

Probability sampling is defined as the kind of sampling in which every element in the population has an equal chance of being selected. The possible inclusion of each population element in this kind of sampling takes place by chance and is attained through random selection. When probability sampling is used, inferential statistics enable researchers to estimate the extent to which the findings based on the sample are likely to differ from what they would have found by studying the whole population. The four types of probability sampling most frequently used in educational research are simple random sampling, stratified sampling, cluster sampling, and systematic sampling.

Simple Random Sampling

The best known of the probability sampling procedures is **simple random sampling.** The basic characteristic of simple random sampling is that all members of the population have an equal and independent chance of being included in the **random sample.** The steps in simple random sampling comprise the following:

1. Define the population.
2. List all members of the population.
3. Select the sample by employing a procedure where sheer chance determines which members on the list are drawn for the sample.

The first step in drawing a random sample from a population is to assign each member of the population a distinct identification number. Let us illustrate this procedure by showing how to obtain a sample of 50 students from the population attending Washington High School. First, you need to enumerate all the individuals in the population. The principal's office could supply a list of all students enrolled in the school. For identification purposes, you would then assign a number to each individual in the population. If there are 800 students in the school, you use the numbers 000, 001, 002, 003, . . . , 799 for this purpose. Each individual must have an identification value with the same number of digits as every other individual. Many schools have already assigned identification numbers to all their students. One way to draw a random sample would be to write the student numbers on separate slips of paper, place the pieces of paper in a container, shake the container, and draw out a slip of paper. Shake the container again, draw out another paper, and continue the process until 50 slips of paper have been picked. This process would be very tedious. A more systematic way to obtain a random sample is to use a table of random numbers, which includes a series of numbers, typically four to six digits in length, arranged in columns and rows (see Table 7.1 for a small segment of a table). A table of random numbers is produced by a computer program that guarantees that all the digits (0–9) have an equal chance of occurring each time a digit is printed. Most statistics books include a table of random numbers in the appendix. In previous editions of this book, we included a five-page table of random numbers. We decided this is no longer needed because there are so many tables available on the Internet, in statistics texts, and from other sources.

Let us illustrate how to use a table of random numbers. With our list of the 800 students in the population, we will use a table to obtain numbers of three digits each, using only those numbers that are less than or equal to 799. For each number chosen, the corresponding member of the population falls into the sample. Continue the process until the desired number for the sample has been chosen—in this case, the first 50 numbers that meet the criterion.

We begin by randomly selecting a starting point in the table. You can do this by closing your eyes and putting your finger on the page, or you can use a procedure that is an absolutely random way to enter the table. First, roll a die to determine which page to use. We roll a 3, so we pull the third page from a table of random numbers (Table 7.1). Then we note the last two digits from the serial number on a dollar bill. They are 03, so we go to row 3. Then we take the last two digits from a second dollar bill, which are 22, taking us to the intersection of row 3 and column 22. The intersection of the row and column is the location of the first random number. Because our population is 800, we will only look at the first three digits of the numbers in the table. If the population were 1500, we would look at the first four digits. In our example, we could use either the first three digits or the last three; we have chosen to use the first three. The first three digits from that intersection are 403, so the individual with number 403 is in the sample. Because the digits in a table are random, the numbers can be used vertically in both directions or horizontally in both directions. You should specify the direction you will use prior to entering the table and use it consistently. The remaining numbers would be located by moving in the specified direction. If we have decided to move vertically, the next three digits are 497, 243, 262, 782, and on down the column through 351. The next number is 995, which is larger than 799 (the size of the sample) so we would skip it and move on down, selecting the numbers smaller than 799. We have highlighted the numbers in that column that would be selected. You would then move to the next column and continue the process until you have 50 random numbers less than 799.

You probably will not actually have to do all this. However, we wanted to show you a way in which the numbers drawn from a table of random numbers can be absolutely without bias. You most likely will have access to web-based random number generators such as Research Randomizer (www.randomizer.org). If you access this website, you will find information about the Research Randomizer and a tutorial on how to use it to generate random numbers quickly. It is part of the Social Psychology Network and is free. Or, you may be lucky and conduct your research in a school whose record-keeping system allows for drawing a random sample using the school's computer.

The generally understood meaning of the word *random* is "without purpose or by accident." However, random sampling is purposeful and methodical. It is apparent that a sample selected randomly is not subject to the biases of the researcher. Rather, researchers commit themselves to selecting a sample in such a way that their biases are not permitted to operate; chance alone determines which elements in the population will be in the sample. They are pledging to avoid a deliberate selection of subjects who will confirm the hypothesis.

Table 7.1 Page from a Table of Random Numbers								
	Column Number							
	00000	00000	11111	11111	22222	22222	33333	33333
Row	01234	56789	01234	56789	01234	56789	01234	56789
				3rd Thousand				
00	89221	02362	65787	74733	51272	30213	92441	39651
01	04005	99818	63918	29032	94012	42363	01261	10650
02	98546	38066	50856	75045	40645	22841	53254	44125
03	41719	84401	59226	01314	54581	40398	49988	65579
04	28733	72489	00785	25843	24613	49797	85567	84471
05	65213	83927	77762	03086	80742	24395	68476	83792
06	65553	12678	90906	90466	43670	26217	69900	31205
07	05668	69080	73029	85746	58332	78231	45986	92998
08	39202	99718	49757	79519	27387	76373	47262	91612
09	64592	32254	45879	29431	38320	05981	18067	87137
10	07513	48792	47314	83660	68907	05336	82579	91582
11	86593	68501	56638	99800	82839	35148	56541	07232
12	83735	22599	97977	81248	36838	99560	32410	67614
13	08595	21826	54655	08204	87990	17033	56258	05384
14	41273	27149	44293	69458	16828	63962	15864	35431
15	00473	75908	56238	12242	72631	76314	47252	06347
16	86131	53789	81383	07868	89132	96182	07009	86432
17	33849	78359	08402	03586	03176	88663	08018	22546
18	61870	41657	07468	08612	98083	97349	20775	45091
19	43898	65923	25078	86129	78491	97653	91500	80786
20	29939	39123	04548	45985	60952	06641	28726	46473
21	38505	85555	14388	55077	18657	94887	67831	70819
22	31824	38431	67125	25511	72044	11562	53279	82268
23	91430	03767	13561	15597	06750	92552	02391	38753
24	38635	68976	25498	97526	96458	03805	04116	63514

You would expect a random sample to be representative of the target population sampled. However, a random selection, especially with small samples, does not absolutely guarantee a sample that will represent the population well. Random selection does guarantee that any differences between the sample and the parent population are only a function of chance and not a result of the researcher's bias. The differences between random samples and their parent population are not systematic. For example, the mean reading achievement of a random sample of sixth-graders may be higher than the mean reading achievement of the target population, but it is equally likely that the mean for the sample will be lower than the mean for the target population. In other words, with random sampling the sampling errors are just as likely to be negative as they are to be positive.

Furthermore, statistical theorists have shown, through deductive reasoning, how much a researcher can expect the observations derived from random samples to differ from what would be observed in the population when the null hypothesis is true. All inferential statistical procedures have this aim in mind. When random sampling is used, the researcher can employ inferential statistics to estimate how much the population is likely to differ from the sample. The inferential statistics in this chapter are all based on random sampling and apply directly only to those cases in which the sampling has been random.

Unfortunately, simple random sampling requires enumeration of all individuals in a finite population before the sample can be drawn—a requirement that often presents a serious obstacle to the practical use of this method. Now let us look at other probability sampling methods that approximate simple random sampling and may be used as alternatives in certain situations.

Stratified Sampling

When the population consists of a number of subgroups, or strata, that may differ in the characteristics being studied, it is often desirable to use a form of probability sampling called **stratified sampling.** For example, if you were conducting a poll designed to assess opinions on a certain political issue, it might be advisable to subdivide the population into subgroups on the basis of age, neighborhood, and occupation because you would expect opinions to differ systematically among various ages, neighborhoods, and occupational groups. In stratified sampling, you first identify the strata of interest and then randomly draw a specified number of subjects from each stratum. The basis for stratification may be geographic or may involve characteristics of the population such as income, occupation, gender, age, year in college, or teaching level. In studying adolescents, for example, you might be interested not merely in surveying the attitudes of adolescents toward certain phenomena but also in comparing the attitudes of adolescents who reside in small towns with those who live in medium-size and large cities. In such a case, you would divide the adolescent population into three groups based on the size of the towns or cities in which they reside and then randomly select independent samples from each stratum.

An advantage of stratified sampling is that it enables the researcher to also study the differences that might exist between various subgroups of a population. In this kind of sampling, you may either take equal numbers from each stratum or select in proportion to the size of the stratum in the population. The latter procedure is known as **proportional stratified sampling,** which is applied when the characteristics of the entire population are the main concern in the study. Each stratum is represented in the sample in exact proportion to its frequency in the total population. For example, if 10 percent of the voting population are college students, then 10 percent of a sample of voters to be polled would be taken from this stratum. If a superintendent wants to survey the teachers in a school district regarding some policy and believes that teachers at different levels may feel differently, he or she could stratify on teaching level and then select a number from each level in proportion to its size in the total population of teachers. If 43 percent of the teachers are high school teachers, then 43 percent of the sample would be high school teachers.

In some research studies, however, the main concern is with differences among various strata. In these cases, the researcher chooses samples of equal size from each stratum. For example, if you are investigating the difference between the attitudes of graduate and undergraduate students toward an issue, you include equal numbers in both groups and then study the differences that might exist between them. You choose the procedure according to the nature of the research question. If your emphasis is on the types of differences among the strata, you select equal numbers of cases from each. If the characteristics of the entire population are your main concern, proportional sampling is more appropriate. When the population to be sampled is not homogeneous but consists of several subgroups, stratified sampling may give a more representative sample than simple random sampling. In simple random sampling, certain strata may by chance be over- or underrepresented in the sample. For example, in the simple random sample of high school students it would be theoretically possible (although highly unlikely) to obtain female subjects only. This could not happen, however, if males and females were listed separately and a random sample were then chosen from each group. The major advantage of stratified sampling is that it guarantees representation of defined groups in the population.

Cluster Sampling

As mentioned previously, it is very difficult, if not impossible, to list all the members of a target population and select the sample from among them. The population of American high school students, for example, is so large that you cannot list all its members for the purpose of drawing a sample. In addition, it would be very expensive to study a sample that is scattered throughout the United States. In this case, it would be more convenient to study subjects in naturally occurring groups, or clusters. For example, a researcher might choose a number of schools randomly from a list of schools and then include all the students in those schools in the sample. This kind of probability sampling is referred to as **cluster sampling** because the unit chosen is not an individual but, rather, a group of individuals who are naturally together. These individuals constitute a cluster insofar as they are alike with respect to characteristics relevant to the variables of the study. To illustrate, let us assume a public opinion poll is being conducted in Atlanta. The investigator would probably not have access to a list of the entire adult population; thus, it would be impossible to draw a simple random sample. A more feasible approach would involve the selection of a random sample of, for example, 50 blocks from a city map and then the polling of all the adults living on those blocks. Each block represents a cluster of subjects, similar in certain characteristics associated with living in proximity. A common application of cluster sampling in education is the use of intact classrooms as clusters.

It is essential that the clusters actually included in your study be chosen at random from a population of clusters. Another procedural requirement is that once a cluster is selected, *all* the members of the cluster must be included in the sample. The sampling error (discussed later) in a cluster sample is much greater than in true random sampling. It is also important to remember that if the number of clusters is small, the likelihood of sampling error is great—even if the total number of subjects is large.

Systematic Sampling

Still another form of probability sampling is called **systematic sampling.** This procedure involves drawing a sample by taking every *K*th case from a list of the population.

First, you decide how many subjects you want in the sample (n). Because you know the total number of members in the population (N), you simply divide N by n and determine the sampling interval (K) to apply to the list. Select the first member randomly from the first K members of the list and then select every Kth member of the population for the sample. For example, let us assume a total population of 500 subjects and a desired sample size of 50: $K = N/n = 500/50 = 10$.

Start near the top of the list so that the first case can be randomly selected from the first 10 cases, and then select every tenth case thereafter. Suppose the third name or number on the list was the first selected. You would then add the sampling interval, or 10, to 3—and thus the 13th person falls in the sample, as does the 23rd, and so on—and would continue adding the constant sampling interval until you reached the end of the list.

Systematic sampling differs from simple random sampling in that the various choices are not independent. Once the first case is chosen, all subsequent cases to be included in the sample are automatically determined. If the original population list is in random order, systematic sampling would yield a sample that could be statistically considered a reasonable substitute for a random sample. However, if the list is not random, it is possible that every Kth member of the population might have some unique characteristic that would affect the dependent variable of the study and thus yield a biased sample. Systematic sampling from an alphabetical list, for example, would probably not give a representative sample of various national groups because certain national groups tend to cluster under certain letters, and the sampling interval could omit them entirely or at least not include them to an adequate extent.

Note that the various types of probability sampling that have been discussed are not mutually exclusive. Various combinations may be used. For example, you could use cluster sampling if you were studying a very large and widely dispersed population. At the same time, you might be interested in stratifying the sample to answer questions regarding its different strata. In this case, you would stratify the population according to the predetermined criteria and then randomly select the cluster of subjects from among each stratum.

NONPROBABILITY SAMPLING

In many research situations, the enumeration of the population elements—a basic requirement in probability sampling—is difficult, if not impossible. Or a school principal might not permit a researcher to draw a random sample of students for a study but would permit use of certain classes. In these instances, the researcher would use nonprobability sampling, which involves nonrandom procedures for selecting the members of the sample. In nonprobability sampling, there is no assurance that every element in the population has a chance of being included. Its main advantages are convenience and economy. The major forms of nonprobability sampling are convenience sampling, purposive sampling, and quota sampling.

Convenience Sampling

Convenience sampling, which is regarded as the weakest of all sampling procedures, involves using available cases for a study. Interviewing the first individuals you encounter on campus, using a large undergraduate class, using the students in your own classroom as a sample, or taking volunteers to be interviewed in survey research are various examples of convenience sampling. There is no way (except by repeating the study using probability sampling) of estimating the error introduced by the convenience sampling procedures. Probability sampling is the ideal, but in practice, convenience sampling may be all that is available to a researcher. In this case, a convenience sample is perhaps better than nothing at all. If you do use convenience sampling, be extremely cautious in interpreting the findings and know that you cannot generalize the findings.

Purposive Sampling

In **purposive sampling—**also referred to as **judgment sampling—**sample elements judged to be typical, or representative, are chosen from the population. The assumption is that errors of judgment in the selection will counterbalance one another. Researchers often use purposive sampling for forecasting national elections. In each state, they choose a number of small districts whose returns in previous elections have been typical of the entire state. They interview all the eligible voters in these districts and use the results to predict the voting patterns of the state. Using similar procedures in all states, the pollsters forecast the national results.

The critical question in purposive sampling is the extent to which judgment can be relied on to arrive at a typical sample. There is no reason to assume that the units judged to be typical of the population will continue to be typical over a period of time. Consequently, the results of a study using purposive sampling may be misleading. Because of its low cost and convenience, purposive sampling has been useful in attitude and opinion surveys. Be aware of the limitations, however, and use the method with extreme caution.

Quota Sampling

Quota sampling involves selecting typical cases from diverse strata of a population. The quotas are based on known characteristics of the population to which you wish to generalize. Elements are drawn so that the resulting sample is a miniature approximation of the population with respect to the selected characteristics. For example, if census results show that 25 percent of the population of an urban area lives in the suburbs, then 25 percent of the sample should come from the suburbs.

Here are the steps in quota sampling:

1. Determine a number of variables, strongly related to the question under investigation, to be used as bases for stratification. Variables such as gender, age, education, and social class are frequently used.
2. Using census or other available data, determine the size of each segment of the population.
3. Compute quotas for each segment of the population that are proportional to the size of each segment.
4. Select typical cases from each segment, or stratum, of the population to fill the quotas.

The major weakness of quota sampling lies in step 4, the selection of individuals from each stratum. You simply do not know whether the individuals chosen are representative of the given stratum. The selection of elements is likely to be based on accessibility and convenience. If you are selecting 25 percent of the households in the inner city for a survey, you are more likely to go to houses that are attractive rather than dilapidated, to those that are more accessible, to those where people are at home during the day, and so on. Such procedures automatically result in a systematic bias in the sample because certain elements are going to be misrepresented. Furthermore, there is no basis for calculating the error involved in quota sampling.

Despite these shortcomings, researchers have used quota sampling in many projects that might otherwise not have been possible. Many believe that speed of data collection outweighs the disadvantages. Moreover, years of experience with quota samples have made it possible to identify some of the pitfalls and to take steps to avoid them.

RANDOM ASSIGNMENT

We distinguish random sampling from random assignment. **Random assignment** is a procedure used after we have a sample of participants and before we expose them to a treatment. For example, if we wish to compare the effects of two treatments on the same dependent variable, we use random assignment to put our available participants into groups. Random assignment requires a chance procedure such as a table of random numbers to divide the available subjects into groups. Then a chance procedure such as tossing a coin is used to decide which group gets which treatment.

As with random sampling, any bias the researcher has will not influence who gets what treatment, and the groups will be statistically equivalent before treatment. Group 1 may have more highly motivated subjects than group 2, but it is just as likely that group 2 will have more highly motivated subjects than group 1. The same is true of all possible known or unknown variables that might influence the dependent variable. Therefore, the same lawful nature of sampling errors that are true of random sampling are true of random assignment.

THE SIZE OF THE SAMPLE (FUNDAMENTALS)

Laypeople are often inclined to criticize research (especially research whose results they do not like) by saying the sample was too small to justify the researchers' conclusions. How large should a sample be? Other things being equal, a larger sample is more likely to be a good representative of the population than a smaller sample. However, the most important characteristic of a sample is its representativeness, not its size. A random sample of 200 is better than a random sample of 100, but a random sample of 100 is better than a biased sample of 2.5 million.

Size alone will not guarantee accuracy. A sample may be large and still contain a bias. The latter situation is well illustrated by the *Literary Digest* magazine poll of 1936, which predicted the defeat of President Roosevelt. Although the sample included approximately 2.5 million respondents, it was not representative of the voters; thus, the pollsters reached an erroneous conclusion. The bias resulted from selecting respondents for the poll from automobile registrations, telephone directories, and the magazine's subscription lists. These subjects would certainly not represent the total voting population in 1936, when many people could not afford automobiles, telephones, or magazines. Also, because the poll was conducted by mail, the results were biased by differences between those who responded and those who did not. We have since learned that with mailed questionnaires, those who are against the party in power are more likely to return their questionnaires than those who favor the party in power. The researcher must recognize that sample size will not compensate for any bias that faulty sampling techniques may introduce. Representativeness must remain the prime goal in sample selection.

Later in this chapter, we introduce a procedure for determining appropriate sample size, on the basis of how large an effect size is considered meaningful and on statistical considerations. Such procedures, known as **power calculations,** are the best way to determine needed sample sizes.

THE CONCEPT OF SAMPLING ERROR

When an inference is made from a sample to a population, a certain amount of error is involved because even random samples can be expected to vary from one to another. The mean intelligence score of one random sample of fourth-graders will probably differ from the mean intelligence score of another random sample of fourth-graders from the same population. Such differences, called **sampling errors,** result from the fact that the researcher has observed only a sample and not the entire population.

Sampling error is "the difference between a population parameter and a sample statistic." For example, if you know the mean of the entire population (symbolized μ) and also the mean of a random sample (symbolized $\overline{X}$) from that population, the difference between these two ($\overline{X} - \mu$) represents sampling error (symbolized e). Thus, $e = \overline{X} - \mu$. For example, if you know that the mean intelligence score for a population of 10,000 fourth-graders is $\mu = 100$ and a particular random sample of 200 has a mean of $\overline{X} = 99$, then the sampling error is $\overline{X} - \mu = 99 - 100 = -1$. Because we usually depend on sample statistics to estimate population parameters, the notion of how samples are expected to vary from populations is a basic element in inferential statistics. However, instead of trying to determine the discrepancy between a sample statistic and the population parameter (which is not often known), the approach in inferential statistics is to estimate the variability that could be expected in the statistics from a number of different random samples drawn from the same population. Because each of the sample statistics is considered to be an estimate of the same population parameter, any variation among sample statistics must be attributed to sampling error.

The Lawful Nature of Sampling Errors

Given that random samples drawn from the same population will vary from one another, is using a sample to make inferences about a population really any better than just guessing? Yes, it is because sampling errors behave in a lawful and predictable manner. The laws concerning sampling error have been derived through deductive logic and have been confirmed through experience.

Although researchers cannot predict the nature and extent of the error in a single sample, they can predict the nature and extent of sampling errors in general. Let us illustrate with reference to sampling errors connected with the mean.

Sampling Errors of the Mean

Some sampling error can always be expected when a sample mean is used to estimate a population mean μ. Although, in practice, such an estimate is based on a single sample mean, assume that you drew several random samples from the same population and computed a mean for each sample. You would find that these sample means would differ from one another and would also differ from the population mean (if it were known). Statisticians have carefully studied sampling errors of the mean and found that they follow known laws.

1. *The expected mean of sampling errors is zero*. Given an infinite number of random samples drawn from a single population, the positive errors can be expected to balance the negative errors so that the mean of the sampling errors will be zero. For example, if the mean height of a population of college freshmen is 5 feet 9 inches and several random samples are drawn from that population, you would expect some samples to have mean heights greater than 5 feet 9 inches and some to have mean heights less than 5 feet 9 inches. In the long run, however, the positive and negative sampling errors will balance. If you had an infinite number of random samples of the same size, calculated the mean of each of these samples, and then computed the mean of all these means, this mean would be equal to the population mean.

   Because positive errors equal negative errors, a single sample mean is as likely to underestimate a population mean as to overestimate it. Therefore, we can justify stating that a sample mean is an unbiased estimate of the population mean and is a reasonable estimate of the population mean.
2. *Sampling error is an inverse function of sample size*. As the size of a random sample increases, there is less fluctuation from one sample to another in the value of the mean. In other words, as the size of a sample increases, the expected sampling error decreases. Small samples produce more sampling error than large ones. You would expect the means based on samples of 10 to fluctuate a great deal more than the means based on samples of 100. In the height example, it is much more likely that a random sample of 4 will include 3 above-average freshmen and 1 below-average freshman than that a random sample of 40 would include 30 above-average and 10 below-average freshman. As sample size increases, the likelihood that the mean of the sample is near the population mean also increases. There is a mathematical relationship between sample size and sampling error. This relationship has been incorporated into inferential formulas, which we discuss later.
3. *Sampling error is a direct function of the standard deviation of the population*. The more spread, or variation, there is among members of a population, the more spread there will be in sample means. For example, the mean weights of random samples of 25, each selected from a population of professional jockeys, would show relatively less sampling error than the mean weights of samples of 25 selected from a population of schoolteachers. The weights of professional jockeys fall within a narrow range; the weights

of schoolteachers do not. Therefore, for a given sample size, the expected sampling error for teachers' weights would be greater than the expected sampling error for jockeys' weights.

4. *Sampling errors are distributed in a normal or near-normal manner around the expected mean of zero*. Sample means near the population mean will occur more frequently than sample means far from the population mean. As you move farther and farther from the population mean, you find fewer and fewer sample means occurring. Both theory and experience have shown that the means of random samples are distributed in a normal or near-normal manner around the population mean. Because a sampling error in this case is the difference between a sample mean and the population mean, the distribution of sampling errors is also normal or near normal in shape.

The distribution of sample means will resemble a normal curve even when the population from which the samples are drawn is not normally distributed. For example, in a typical elementary school you will find approximately equal numbers of children of various ages included, so a polygon of the children's ages would be basically rectangular. If you took random samples of 40 each from a school with equal numbers of children aged 6 through 11 years, you would find many samples with a mean age near the population mean of 8.5 years, sample means of approximately 8 or 9 would be less common, and sample means as low as 7 or as high as 10 would be rare. Note that the word error in this context does not mean "mistake"—it refers to what is unaccounted for.

Standard Error of the Mean

Because the extent and the distribution of sampling errors can be predicted, researchers can use sample means with predictable confidence to make inferences concerning population means. However, you need an estimate of the magnitude of the sampling error associated with the sample mean when using it as an estimate of the population mean. An important tool for this purpose is the standard error of the mean. Sampling error manifests itself in the variability of sample means. Thus, if you calculate the standard deviation of a collection of means of random samples from a single population, you would have an estimate of the amount of sampling error. It is possible, however, to obtain this estimate on the basis of only one sample. We have noted that two things affect the size of sampling error: the size of the sample and the standard deviation in the population. When both of these are known, you can predict the standard deviation of sampling errors. This expected standard deviation of sampling errors of the mean is called the **standard error of the mean** and is represented by the symbol $\sigma_{\overline{X}}$. Deductive logic shows that the standard error of the mean is equal to the standard deviation of the population (σ) divided by the square root of the number in each sample ($\sqrt{n}$). As a formula,

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} \tag{7.1}$$

where

$\sigma_{\overline{X}}$ = standard error of the mean
σ = standard deviation of the population
n = number in each sample

In Chapter 6, we noted that standard deviation (σ) is an index of the degree of spread among individuals in a population. In the same way, standard error of the mean ($\sigma_{\overline{X}}$) is an index of the spread expected among the means of samples drawn randomly from a population. As you will see, the interpretation of σ and $\sigma_{\overline{X}}$ is very similar.

Because the means of random samples have approximately normal distributions, you can also use the normal curve model to make inferences concerning population means. Given that the expected mean of sample means is equal to the population mean, that the standard deviation of these means is equal to the standard error of the mean, and that the means of random samples are distributed normally, you can compute a ***z* score** for a sample mean and refer that z to the normal curve table to approximate the probability of a sample mean occurring through chance that far or farther from the population mean. The z is derived by subtracting the population mean from the sample mean and then dividing this difference by the standard error of the mean:

$$z = \frac{\overline{X} - \mu}{\sigma_{\bar{x}}} \tag{7.2}$$

To illustrate, consider a college admissions officer who wonders if her population of applicants is above average on the verbal subtest of the College Board examination. The national mean for College Board verbal scores is 500, and the standard deviation is 100. She pulls a random sample of 64 from her population and finds the mean of the sample to be 530. She asks the question, How probable is it that a random sample of 64 with a mean of 530 would be drawn from a population with a mean of 500? Using Formula 7.1, the admissions officer calculates the standard error of the mean as 12.5:

$$\begin{aligned} \sigma_{\bar{x}} &= \frac{\sigma}{\sqrt{n}} \\ &= \frac{100}{\sqrt{64}} \\ &= 12.5 \end{aligned}$$

Calculating the z score for her sample mean with Formula 7.2, she obtains the following result:

$$\begin{aligned} z &= \frac{\overline{X} - \mu}{\sigma_{\overline{X}}} \\ &= \frac{530 - 500}{12.5} \\ &= 2.4 \end{aligned}$$

Thus, the sample mean deviates from the population mean by 2.4 standard error units. What is the probability of having a sample mean that deviates by this amount (2.4 $\sigma_{\overline{X}}$) $\overline{X}$ or more from the population mean? It is only necessary to refer to the normal curve table in order to express this deviation (z) in terms of probability. Referring to the normal curve table, the admissions officer finds that the probability of a $z = 2.4$ or higher is .0082. This means that a z score that great or greater would occur by chance only approximately 8 times in 1000. Because the probability of getting a sample mean that far from the population mean is remote, she concludes that the sample mean probably did not come from a population with a mean of 500, and therefore the mean of her population—applicants to her college—is very probably greater than 500.