cl. 7

FUNDAMENTAL STATISTICS FOR BEHAVIORAL SCIENCES

FIFTH EDITION

Robert B. McCall
University of Pittsburgh

Under the General Editorship of
Jerome Kagan
Harvard University

HBJ

Harcourt Brace Jovanovich, Publishers
San Diego New York Chicago Austin Washington, D.C.
London Sydney Tokyo Toronto

COPYRIGHTS AND ACKNOWLEDGMENTS

LONGMAN GROUP Tables B, C, G, and K are taken from Tables III, VII, IV, and XXXIII of Fisher and Yates: *Statistical Tables for Biological, Agricultural and Medical Research* published by Longman Group Ltd., London (previously published by Oliver and Boyd Ltd., Edinburgh) and by permission of the authors and publishers.

ISBN: 0-15-529476-8

Library of Congress Catalog Card Number: 89-85302

Printed in the United States of America

CHAPTER 7

SAMPLING, SAMPLING DISTRIBUTIONS, AND PROBABILITY

THE FIRST SIX CHAPTERS OF THIS BOOK have been concerned with **descriptive statistics**—that is, procedures that describe and summarize groups of measurements. Attention is now turned to **inferential statistics**, which includes techniques for making inferential decisions when only partial information is available.

Statistical inference consists of using probability to make decisions about a population on the basis of a sample of observations. In this chapter we will discuss a number of issues related to this method, including (1) the adequacy of the sample as a faithful representation of the population, (2) the sampling distribution as an index of the amount of sampling error, and (3) how the relative frequency of sampling distributions can be used to make probability statements about the likelihood of certain events.

METHODS OF SAMPLING

Recall from Chapter 3 that a **population** is a complete set of subjects, events, or scores that have some common characteristic. Characteristics of populations are known as **parameters**. A subset or portion of a population is a **sample**, and quantities computed on a sample are called **statistics**. Often a statistic is used to estimate the value of a parameter. For example, a pollster might select a sample of 1000 cases with which to estimate the percentage of the American adult population that supports the president. A statistic used to estimate a parameter is no better than the sample upon which it is computed.

Simple Random Sampling

There are many different ways of selecting a sample, but one of the most commonly used methods is simple random sampling.

In simple random sampling, all elements of the population have an equal probability of being selected for the sample.

Suppose you want to have a random sample of students at your university for an opinion poll on the quality and appropriateness of their educational experience. You might obtain a list of all students in the school, go to a *random number table*, and select a sample of 50. A random number table (Table K) is provided in Appendix 2 of this book. It consists of rows and columns of random numbers. The numbers are random in the sense that for any single digit position, each of the 10 numbers from 0 to 9 has an equal chance of occupying that position. This means that every digit was selected independently of every other digit. Further, not only are all the single digits random, but all two-, three-, or *N*-digit numbers are also random.

To use the table for getting a sample of 50 students from a list of 3000, assign each student a number between 1 and 3000. Then go to the random number table and mentally block the table off into four-digit columns. Read down the columns until you obtain 50 four-digit numbers that fall between 0001 and 3000 inclusive. The students assigned numbers corresponding to these 50 numbers constitute a randomly selected sample.

If you read down a one-digit column in the table, you may feel that the numbers are not random at all. For example, if you were to write what you thought was a random sequence of 0s and 1s, you might write 01001001011101, and so on. Most people are very hesitant to put more than three of a kind in succession (for example, 0000 or 1111). Yet, in any random selection of four 0s or 1s, the probability of 0000 or 1111 is .125 or one in every eight such sets of four digits. Look at the first column of numbers in the random number table and observe such a statistically random sequence by letting even numbers be called 0 and odd numbers be called 1. The point of this explanation is that random sampling refers to the process of *how* a sample is selected, not to the cases actually selected. A sample selected by a random process nevertheless might not be very average or typical because atypical samples have a chance of being selected occasionally too.

Random samples tend to have two other important characteristics of sampling in addition to randomness. In the long run, large random samples will be **representative samples** of all aspects of the population. This means that if a population contains 10% Catholics, a random sample will tend toward having 10% Catholics as the sample size increases. But a small, randomly collected sample will not necessarily have 10% Catholics. To ensure that 10% of a sample is Catholic, a random sample may be taken of all Catholics in the population until their total number represents exactly 10% of the intended sample size. Such a sample is called a **proportional stratified random sample**. Political pollsters, for example, attempt to sample the voting public in such a way as to ensure that each area of the country, each ethnic group, each religious group, and so forth, is appropriately represented. The sampling is random and independent within these groups, but not between them. Since all of the procedures outlined in this text are appropriate for simple random samples, stratified sampling and its associated statistical procedures will not be considered further.

In most cases, random sampling also means that each subject is selected independently of other subjects. **Independence** in sampling implies that the selection of any one element, or subject, for inclusion in the sample does not alter the likelihood of drawing any other element of the population into the sample. In almost all cases in which random sampling is required in this text, the implication is that the elements have been independently selected.

Sampling in Practice

In actual practice it is extremely difficult to obtain a truly random sample. Let's suppose, for example, that one wanted to sample the population of a given town by selecting every two-hundredth name in the telephone book. Would it be possible in this way to obtain a random sample of the townsfolk —that is, would each person in the town have an equal likelihood of being selected? The answer is of course no, since people who do not own a phone and those who have unlisted phone numbers would automatically be excluded from the selection process. There are two things one could do to correct this situation. First, a list of all people in the town could be obtained from the city government or census bureau, and a random number table could then be employed. But this might be a very time-consuming task, and it assumes the availability of the necessary list. The second possibility is to change the definition of the population in accordance with the nature of the sample. Since you cannot get a random sample of all the townspeople, you could change the population to include only those listed in the telephone book and then discuss your results and conclusions in terms of this group rather than in terms of all residents. The important point is for the researcher to be aware of precisely what population the sample was selected from and to limit conclusions to that population.

Even when scientists have control over their subjects, such as when rats are used, samples are not truly random. For example, a group of rats provided by an animal supplier is not a random selection of rats. Rats are raised in cages set on tiers, some of which are closer to the light than others, and the amount of illumination in the rat's rearing experience can influence some types of later behavior. Further, when the rats arrive at the laboratory and are assigned to experimental groups, it is sometimes tempting to place the first 10 in one group, the second 10 in the next, and so on. But it happens that the more curious and active rats frequently come over to the side of the shipping box when it is opened. These rats are more accessible and easier to pick up. They are thus selected first and go into the first group if the above procedure is carried out. It is a good idea to use a random number table to determine group assignments; in any case, one should at least alternate the rats by assigning one to the first group, another to the second, and so on.

Another form of bias in sampling occurs when human volunteers are used. College students who volunteer for experiments are probably somewhat different with respect to academic ability, motivation, and so on from students who do not volunteer. Who volunteers can be an interesting issue on its own right, and it determines what kind of research can and cannot be done. For example, some developmental psychologists depend upon parents to volunteer their infants for observation. It is likely that highly educated parents are more receptive to science's "need" for subjects than are less educated parents. Thus, the sample one obtains is not random because it may contain a preponderance of infants from highly educated families.

How can you safeguard against such bias in your sample? The best way is simply to be cautious and aware of what you are doing. In addition, it is advisable to measure the sample you have selected on several dimensions (such as, age, education, "normality," and so on) appropriate to the research (as long as the measurement of these traits does not influence the subjects in any way) so that readers can judge whether the sample has the general characteristics of the population being discussed. The goal is to ensure that inferences and generalizations are made to the appropriate population. Further, while the statistical procedures in this book require independent random sampling, they are used even when the samples are *not* random or independent. The careful researcher does the best job possible and notes when procedures are likely to violate these assumptions and produce bias in the results.

SAMPLING DISTRIBUTIONS AND SAMPLING ERROR

Even when the sample is random and appropriate, it is likely to be different from the population because it is a sample of considerably fewer cases than are contained in the population. In short, as a method of estimating the population, sampling is an imprecise process.

An Empirical Sampling Distribution

Suppose there are 20 students in your statistics class and the professor springs a surprise quiz of 10 questions. The scores of the 20 students comprise the population of raw scores and are presented in the left-hand column of Table 7–1. Now suppose a random sample of $N = 4$ students is taken for the purpose of estimating the population mean (the mean of the 20 students, which is 3.90). The first such sample of four students provides the scores (1, 5, 9, 0), which have a mean of 3.75, as indicated in Table 7–1. But suppose another random sample of $N = 4$ is selected and its mean is 2.25. Looking down the right-hand column in Table 7–1, which shows the means for 10 randomly selected samples from the population, one can see that randomly selected samples from the same population vary in the value of their means. Moreover, not one sample yields a mean that precisely equals the population mean. Why do samples differ from another?

In a random sample, each subject in the population has an equal *opportunity* to be drawn into the sample. But this is not to say that the sample that is actually drawn will faithfully reflect the population's characteristics.

7–1 POPULATION DISTRIBUTION OF 20 RAW SCORES, 10 OBSERVED SAMPLE DISTRIBUTIONS (N = 4), AND AN EMPIRICAL SAMPLING DISTRIBUTION OF THE MEANS

Population Distribution of Raw Scores		10 Observed Sample Distributions ($N = 4$)	Empirical Sampling Distribution of the $\overline{X}$'s
6	2	(1, 5, 9, 0)	3.75
9	5	(0, 3, 1, 5)	2.25
0	1	(5, 8, 3, 0)	4.00
3	2	(1, 5, 0, 7)	3.25
1	1	(7, 6, 1, 3)	4.25
5	2	(3, 2, 1, 7)	3.25
7	7	(2, 0, 3, 5)	2.50
7	8	(1, 2, 1, 1)	1.25
1	1	(2, 7, 1, 7)	4.25
3	7	(9, 7, 6, 2)	6.00
$\mu = 3.90$, $\sigma = 2.88$			Mean of $\overline{X}$'s = $\overline{X}_{\bar{x}} = 3.48$ Standard deviation of $\overline{X}$'s = $s_{\bar{x}} = 1.31$

For example, one might randomly select a sample that just happened to contain many exceptionally bright students. One obvious reason why random samples differ from one another is that they are composed of different individuals.

Returning to Table 7–1, we see that the means computed on the 10 samples could themselves be considered scores in a distribution—a distribution of means rather than of raw scores. Such a distribution has a special name and function in statistics.

> **The distribution of a statistic determined on separate independent samples of size N drawn from a given population is called a sampling distribution.**

Thus, the distribution of the 10 sample means in Table 7–1 is a sampling distribution.

The sampling distribution should be carefully distinguished from the two raw-score distributions introduced earlier. The type of distribution of central concern throughout Part 1 of this text is the **sample distribution**, which is a collection of scores obtained from a subgroup of a population. The middle column of Table 7–1 shows 10 sample distributions. Also introduced in Part 1 (Chapter 3) was the concept of a **population distribution**, which is the full array of raw scores that includes the sample distribution. Now we have another type of distribution. A **sampling distribution** is a distribution of a statistic, not of raw scores. Sampling distributions may be of two general types. The right-hand column of Table 7–1 shows an empirical sampling distribution. The word *empirical* signifies "experienced" or "observed," and these 10 means are observations presumably made by actually collecting 10 samples and computing $\overline{X}$ for each group. In contrast, a theoretical sampling distribution is a theoretical distribution of a statistic, and its characteristics are determined mathematically rather than by repeated observations.

Note that a sampling distribution differs from both a sample distribution and a population distribution in that it is a collection of statistics rather than a collection of raw scores. The statistic whose sampling distribution is shown in the third column of Table 7–1 is the mean, but one can also have sampling distributions of other statistics. For example, one could compute the standard deviations of the ten sample distributions shown in the table and thus obtain a sampling distribution of the standard deviation.

Just as a distribution of raw scores has certain characteristics, such as a mean and a standard deviation, so too does the sampling distribution of a

7–2 TERMS AND SYMBOLS FOR THE MEAN, STANDARD DEVIATION, AND VARIANCE IN DIFFERENT TYPES OF DISTRIBUTIONS

Distributions of Raw Scores	**Mean**	**Standard Deviation**	**Variance**
Sample	$\overline{X}$	s_x or s	s_x^2 or s^2
Population	μ_x or μ	σ_x or σ	σ_x^2 or σ^2
Sampling Distribution of the Mean	**Mean**	**Standard Error of the Mean**	**Square of the Standard Error of the Mean**
Sample	$\overline{X}_{\bar{x}}$	$s_{\bar{x}}$	$s_{\bar{x}}^2$
Population	$\mu_{\bar{x}}$ or μ	$\sigma_{\bar{x}}$	$\sigma_{\bar{x}}^2$

Equivalences

$$\mu_x = \mu_{\bar{x}} = \mu \qquad s_{\bar{x}} = \frac{s_x}{\sqrt{N}} \text{ and } \sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{N}}$$

statistic. The sampling distribution shown in Table 7–1, for example, itself has a mean of 3.48 and a standard deviation of 1.31. Therefore, it will be necessary to have terms and symbols to represent the mean, standard deviation, and variance of the several distributions distinguished above. Table 7–2 summarizes these terms and symbols. First, notice at the left that distributions are composed either of raw scores or of statistics, and in the latter case they are called sampling distributions. In addition, a distribution may be based on a sample or a population. As mentioned above, distributions are also either empirical (that is, based on actual observations) or theoretical (that is, defined in terms of their characteristics—mean, variance, shape—and not in terms of actual observations). In practice, samples are usually empirical, and populations are usually theoretical, but this is not always the case; for example, see the empirical population in Table 7–1. Similarly, in practice, sampling distributions are almost always theoretical, but Table 7–1 contains an empirical sampling distribution to illustrate the concept.

Each of the four kinds of distributions listed in Table 7–2 has a mean, standard deviation, and variance, but they have different names and symbols depending on the type of distribution to which they pertain. When these quantities refer to samples or empirical distributions, they are symbolized by the letters $\bar{X}$, s, and s^2; when they pertain to populations or theoretical distributions, they are symbolized by the Greek letters μ, σ, and σ^2. A subscript designates the distribution to which the quantity refers.

It is important for students, when reading the remainder of this chapter and the next chapter, to pay special attention to the modifiers of particular statistics and distributions. One must be careful to note whether it is the *sample* mean or the *population* mean that is being discussed, or the *sample* standard deviation or the *population* standard deviation. Also, note whether it is a *sample* distribution, a distribution of *raw scores* (the same as a *sample* distribution), a *sampling* distribution, a distribution of the *means* (the same as the *sampling* distribution of the mean), or a *theoretical sampling* distribution. Some students find it helpful to read the remaining material more slowly to ensure that they are thinking about the right concept.

Sampling Statistics

Statistics may pertain to samples or to sampling distributions. **Sample statistics** are quantities that characterize samples of raw scores. While they are calculated on samples of raw scores, they often are used to estimate characteristics of a population of raw scores. **Sampling statistics** are quantities that characterize sampling distributions of statistics. Typically, they are calculated on a single sample of raw scores but are used to estimate the characteristics of a theoretical sampling distribution of a statistic.

In this chapter we focus on two sampling statistics—the mean and standard deviation of the sampling distribution of means. We consider their special names, how they are related to their corresponding statistics based on a population of raw scores rather than a population of sample means, and how their values may be estimated from a single sample of cases. The material presented below is summarized in Table 7–2.

The first concept is the mean of the population of sample means, symbolized by $\mu_{\bar{X}}$. The Greek μ indicates that this is a parameter, not a statistic, and the subscript $\bar{X}$ signifies that it is based on the population or theoretical distribution of sample means.

The population mean of the sampling distribution of means, $\mu_{\bar{X}}$, can be distinguished from the population mean of the raw scores, μ_X, by the different subscripts. However, it happens that these two population values are identical. Symbolically,

$$\mu_{\bar{X}} = \mu_X = \mu$$

That is, the mean of the population sampling distribution of the means ($\mu_{\bar{X}}$) equals the mean of the population of raw scores (μ_X), and the symbol μ without a subscript is customarily used to indicate this value. It is called simply "the population mean" or "mu" after its Greek symbol.

The value of the population mean may be estimated by the mean of a sample of raw scores. That is, $\bar{X}$ estimates μ. Thus, the population mean of raw scores, μ_X, and the population mean of the sampling distribution of means, $\mu_{\bar{X}}$, can both be estimated by drawing a sample of raw scores and calculating the mean, $\bar{X}$.

Consequently, the population mean of the sampling distribution of means, $\mu_{\bar{X}}$, can be estimated without drawing several samples and calculating the mean over all such samples as was done above in Table 7–1 when an empirical sampling distribution was created. The mean of a single sample of raw scores, $\bar{X}$, can be used to estimate $\mu_{\bar{X}}$, μ_X, and μ. Of course, it is not a perfect estimate, as discussed below.

Besides a mean, a sampling distribution has a standard deviation.

The standard deviation of a sampling distribution of a statistic is called the standard error of that statistic. Consequently, the standard deviation of the sampling distribution of the mean is known as the standard error of the mean. In the population, it is symbolized by $\sigma_{\bar{X}}$.

The standard error of the mean is simply the standard deviation of the sampling distribution of means. Earlier, in Table 7–1, an *empirical* standard error of the mean was actually calculated for the 10 samples and found to be 1.31. In the population or in the theoretical sampling distribution of the mean, however, the standard error of the mean, symbolized by $\sigma_{\bar{X}}$, is not calculated.

Again, the Greek letter σ indicates that this is a population or theoretical quantity, not a sample value, and the subscript $\bar{X}$ indicates that it is the standard deviation (or standard error) of the sampling distribution of means. The $\sigma_{\bar{X}}$ can be distinguished from the population standard deviation of raw scores, σ_X, by the different subscripts.

Because of the great importance of the concept of a standard error of the mean in inferential statistics, it is crucial for you to have a firm grasp of its meaning. We have said that the mean of one sample of scores will not likely equal the mean of another sample of scores, even if both samples are randomly selected from the same population. The standard deviation is a numerical index of variability in such a distribution of means. Therefore, the standard deviation of the sampling distribution of means, $\sigma_{\bar{X}}$, is a numerical index of the extent to which means vary from one sample to another.

More generally, the standard error of the mean is an index of the amount of error that results when a single sample mean is used to estimate the population mean; that is, it is an index of **sampling error**. For example, if $\sigma_{\bar{X}} = 5$ for samples of 20 males on a reading test but $\sigma_{\bar{X}} = 10$ for samples of 20 females on the same test, there is less sampling error for males than for females. This implies that the more or less random variation between means from one sample to another is less for males than for females.

It was stated above that the mean of the theoretical sampling distribution of the mean is identical to the mean of the population distribution of raw scores, $\mu_{\bar{X}} = \mu_X = \mu$. In contrast, the standard deviation of the sampling distribution of the mean (the standard error of the mean), $\sigma_{\bar{X}}$, is *not* identical to the standard deviation of the population distribution of raw scores, σ_X, but it is related to it.

The standard error of the mean, $\sigma_{\bar{x}}$, equals the standard deviation of the population of raw scores divided by the square root of the size of the sample on which the means are based:

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{N}}$$

If the standard deviation of the population of raw scores is $\sigma_x = 2.88$ and the sample size is $N = 4$, the theoretical standard error of the mean is

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{N}} = \frac{2.88}{\sqrt{4}} = 1.44$$

How can the value of $\sigma_{\bar{X}}$ be estimated? Just as a sample mean, $\bar{X}$, could be used to estimate μ, the sample standard deviation, s_x, can be used to calculate $s_{\bar{x}}$ which can be used to estimate $\sigma_{\bar{X}}$.

The population standard error of the mean, $\sigma_{\bar{x}}$, may be estimated by

$$s_{\bar{x}} = \frac{s_x}{\sqrt{N}}$$

in which N is the size of the sample of X's.

Two things should be noticed about the standard error of the mean as expressed by $s_{\bar{x}} = s_x/\sqrt{N}$. First, all that is required to calculate $s_{\bar{x}}$ is the standard deviation and the N from a single sample of cases, yet $s_{\bar{x}}$ represents an estimate of the amount of variability (or sampling error) in means from *all* possible samples of size N from the population of raw scores. Thus, it is not necessary to select several samples in order to estimate the population sampling error of the mean; $s_{\bar{x}}$ estimates $\sigma_{\bar{x}}$, and all that is required to calculate $s_{\bar{x}}$ is s_x and N from a single sample of raw scores.

Second, observe that the formula for $s_{\bar{x}}$ states that the standard deviation of the sample must be divided by the square root of N; that is, $s_{\bar{x}} = s_x/\sqrt{N}$. Therefore, the variability of means from sample to sample will always be smaller than the variability of raw scores. Also note that as N becomes larger, $s_{\bar{x}}$ becomes smaller. Thus, the variability of sample means decreases as the size of the sample increases. Consequently, for large samples one expects $\bar{X}$, the sample estimator of the population mean μ, to be less variable from sample to sample, and thus a more accurate estimate of μ, than if the sample size were smaller. In short, when parameters must be estimated, it is a good idea to have as large a sample as possible.

Other Standard Errors

The sampling distribution and standard error of the mean have been discussed in detail, but a sampling distribution and a standard error exist for any statistic. In each case the logic is the same. Random samples differ in their characteristics, and any statistic will vary somewhat from sample to sample. The theoretical sampling distribution is the distribution of a particular statistic determined for all possible samples of size N, and the standard deviation of that statistic's sampling distribution is its standard error. Therefore, one can imagine standard errors for the median, the variance, and even for the difference between two sample means. In each case the standard error reflects the relative extent of the error in using that sample statistic to estimate its corresponding population parameter.

Sampling Distributions and Normality

Since statistics calculated on a single sample may be used to estimate parameters of sampling distributions, it is never necessary to actually collect an *empirical* sampling distribution. Empirical sampling distributions are not used in statistics except to help students understand the concept of a distribution of a statistic. From this point forward, *sampling distribution* will refer to a *theoretical sampling distribution*. However, the symbol $s_{\bar{x}}$ will still be used to indicate that $\sigma_{\bar{x}}$ is being estimated by sample observations.

Many of the procedures described in this and later chapters rest on the assumption that the sampling distribution of means is normal in form. This is the case if either of two conditions is met.

Given random sampling, the sampling distribution of the mean

1. **is a normal distribution if the population distribution of the raw scores is normal, and**
2. **approaches a normal distribution as the size of the sample increases even if the population distribution of raw scores is not normal.**

If the population distribution of raw scores is normal, the sampling distribution of the mean will also be normal. However, since the population is rarely available, how can you know if the population distribution is normal? One way to make an educated guess is to determine whether a random sample from the population is normally distributed. Alternatively, some variables are known to be normally distributed in certain groups. Height or IQ among 21-year-old males, for example, are likely to be normally distributed. But some variables are usually not normally distributed. For example, while the IQs of all nonretarded 21-year-olds are probably normally distributed, the IQs of all 21-year-old college students are not because low scores are not represented as frequently in college groups as are extremely high scores. Family income, the latency for a rat to move out of a startbox in a maze, and percentage correct on a relatively easy exam are variables that are not usually normally distributed. Notice that these variables are bounded on one end of their scales (for example, $0 income, 0 seconds, 100% correct). The scores will tend to fall near the bounded end of the scale, and the distribution is likely to be skewed toward the other direction. Fortunately, many of the variables measured in social sciences can be assumed to be normally distributed. When variables are not normal, the statistical techniques described in Chapter 14, rather than those described below, may be used.

A second way to obtain a normal sampling distribution of the mean is to select a large enough sample of raw scores. The sampling distribution of the mean will approach a normal distribution as the sample size increases, *even though the population distribution of raw scores on the variable of interest is not normal.* Just how many cases constitute a sufficiently large sample depends upon many factors, one of which is the extent of the departure from normality of the population distribution. If the population distribution does not deviate too much from normality, samples of size $N = 2$ might produce a sampling distribution of the mean that is quite normal, whereas if the nonnormality in the population is severe, N's of 20, 30, or several hundred might be necessary. In short, the sampling distribution of the mean will approach a normal form as the size of the samples of raw scores increases. This crucial principle of statistics is called the *Central Limit Theorem.*

The following fact is one of the reasons normality is necessary for the statistical procedures to be described:

If the population distribution of raw scores is normal and the observations are independent and randomly selected, the sample mean and variance (or standard deviation) are independent of one another across samples.

Any two variables (including the statistics $\overline{X}$ and s^2) are independent if they are unrelated to each other. This means that the value of one tells you nothing about the value of the other, and increasing or decreasing the value of one does not necessarily change the value of the other. The independence of the mean and variance of a normal distribution, for example, will be important later.

To review, most of the statistical procedures to be described subsequently depend on several principles:

1. samples are composed of randomly and independently selected observations;
2. a sampling distribution is the distribution of a statistic; the standard deviation (standard error) of a sampling distribution is an index of the

extent to which the statistic varies from one sample to another (that is, of sampling error);

3. the sampling distribution of the mean is normal if the population distribution of raw scores is normal or the size of the sample is large; and

4. the mean and standard deviation of a random sample from a normal population are independent.

PROBABILITY AND ITS APPLICATION TO HYPOTHESIS TESTING

The purpose of inferential statistics is to assist in making inferences and judgments about what exists on the basis of only partial evidence. This is accomplished by using probability. In a way, most people use subjective probability every day. You want to go to the football game this afternoon, but someone has warned you that it is going to rain. You look outside an hour before the game and, while there are clouds, the sky is not threatening. You make a judgment about going to the game based on your subjective probability that it will not rain.

Scientists, however, prefer to use numerical probability rather than their subjective feelings as an index of the likelihood of events. The numerical probability is public knowledge (all scientists can observe and understand it), the probability of one event can be easily compared with the probability of a different event, and certain rules can be adopted about how high or low the probability must be in order to justify one decision or another. Therefore, a knowledge of the concept of probability is essential to understanding the process of statistical inference and decision making in science. A more thorough examination of probability is offered in Chapter 12.

Probability and Relative Frequency

The determination of a simple numerical probability implies an **idealized experiment**. In figuring the chances of a head or a tail when flipping a coin, one actually assumes an experiment in which the coin is tossed over and over again. It is assumed that both heads and tails are equally likely. On any single throw, either a head or a tail will occur, but in the idealized experiment of repeated flips of a coin, the ratio of heads to all possible outcomes will approach 1/2, or .50. In short, the probability of a head, for example, is the relative frequency of heads versus all other possible outcomes in the idealized experiment.

Probability is theoretical relative frequency — the relative frequency of score values in a theoretical distribution based upon an unlimited number of cases.

From the standpoint of probability, an idealized experiment consists of an unlimited number of cases, and the probability of a particular outcome is the theoretical relative frequency of that outcome in the distribution of all outcomes of the idealized experiment. For example, suppose Figure 7–1 represents the theoretical relative frequency distribution of the IQs of normal 10-year-old American children. This represents the results of our "idealized experiment"—the IQs of the entire population of normal 10-year-old American children are available. The numerical IQ values are the "outcomes" of the idealized experiment. Now consider the probability that a single child selected at random from this population would have an IQ greater than 110. According to the conception of probability discussed above, this probability value should be given by the theoretical relative frequency of scores that exceed 110. The approach to determining the probability here rests on equating the area between the curve and the abscissa with the concept of *theoretical relative*

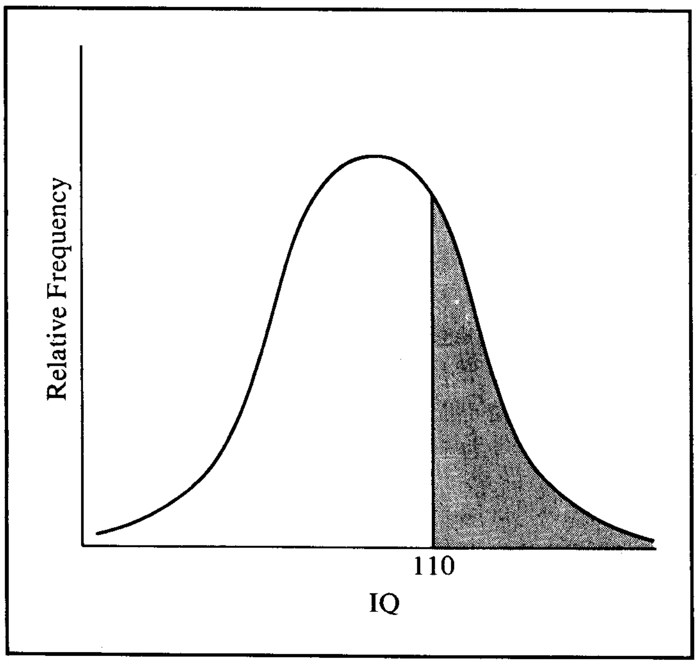

Figure 7–1 **Theoretical relative frequency of the IQs of normal 10-year-old American children. The probability of randomly selecting a child who has an IQ higher than 110 is indicated by the proportion of the total area under the curve represented by the crosshatched area.**

frequency, just as was done in Chapter 4. If the area under the curve in Figure 7–1 represents theoretical relative frequency, then the proportion of the total area that lies between 110 and $+\infty$ (indicated in the figure by the shading) represents the theoretical relative frequency of children having scores greater than 110. The relative frequency of this area is approximately .27, or 27%, of the total. Therefore, the probability of randomly drawing a child from this population having an IQ greater than 110 is approximately .27. To formalize:

The proportion of the total area under the curve of a theoretical relative frequency distribution that exists between any two points represents the probability of obtaining the events contained within the interval delimited by those two points.

Consider another example: what is the probability that a child randomly selected from the population would have an IQ between 85 and 115? This probability is given by the proportion of the total area that exists between IQs of 85 and 115 as illustrated in Figure 7–2. This proportion is approximately 65% of the total, so the relative frequency and probability of selecting a child with an IQ between 85 and 115 is approximately .65.

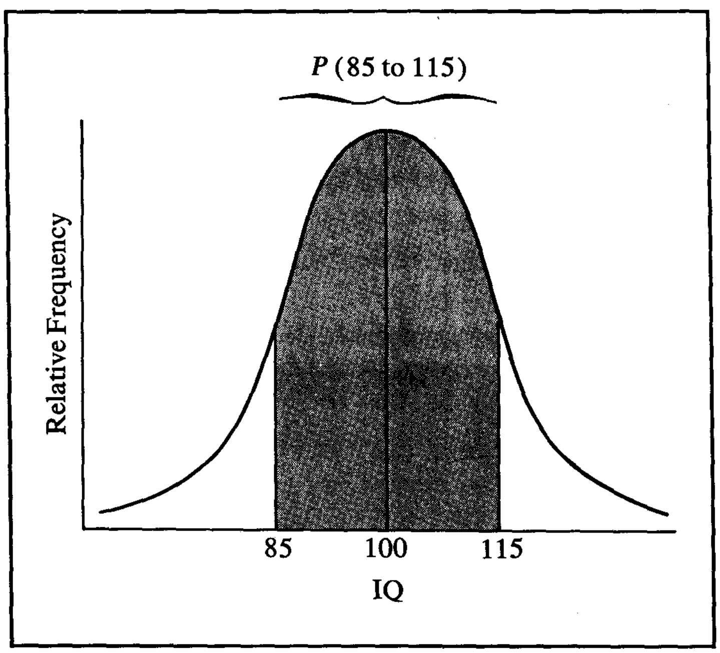

Figure 7–2 **The proportion of the total area under the curve that is crosshatched represents the probability of randomly selecting a child who has an IQ between 85 and 115.**

The Standard Normal Distribution and Probability

The standard normal distribution is a theoretical relative frequency distribution. It can be used in the ways described in Chapter 4 to determine theoretical relative frequency or probability for problems like the two examples just given. The percentiles or relative frequencies for the standard normal distribution are presented in Table A of Appendix 2. In the pages that follow, sampling distributions, which reflect the amount of sampling error in a statistic, will be related to a known theoretical relative frequency distribution,[1] such as the standard normal, and this will permit probabilities to be assigned to the likelihood that one or another circumstance is true. This is the fundamental strategy of inferential statistics.

ESTIMATION[2]

One of the most important purposes of statistical inference is to use sample statistics to estimate their corresponding population parameters. Of course, a statistic must be a "good" estimator of its respective parameter, and you recall from Chapter 3 that the denominator of the formula for the sample variance is $N - 1$ rather than just N to make the sample variance a better estimator of the population variance.

Characteristics of a Good Estimator

A "good" estimator is defined by several criteria, including unbiasedness, consistency, relative efficiency, and sufficiency.

Unbiasedness A good estimator should be unbiased.

> **An unbiased estimator of a population parameter is one whose average over all possible random samples of a given size equals the value of the parameter.**

[1]The equation for a theoretical relative frequency distribution, such as the standard normal and several other distributions to be considered, is known as a **probability**, or **density, function**. To determine the probability of the occurrence of events located between any two points on the dimension, this function is integrated between these points by the methods of calculus. The values in Appendix 2, Table A, and several other tables in this book represent the results of such a process.

[2]The material in the remainder of this chapter is appropriate for more advanced courses and may be omitted.

7–3 AVERAGE OF HYPOTHETICAL SAMPLE MEANS FOR DIFFERENT NUMBERS OF SAMPLES

Number of Samples	Average of Sample Means	Difference between Population Mean (90) and Average Sample Mean
1	73	−17
5	96	+6
10	95	+5
50	87	−3
100	89	−1
1000	90.03	+0.03
∞	90.00	0

In actuality, statisticians determine whether a statistic is an unbiased estimator by using special mathematical procedures called **expectation theory**, which are beyond the scope of this text. But we can get an idea of what the criterion of unbiasedness implies when applied to the sample mean as an unbiased estimator of the population mean by looking at Table 7–3.

Suppose a population has a mean of 90 on a given test. If a single sample of size N were drawn, it is unlikely that the mean of the sample would be exactly 90. Perhaps, for example, it is 73, a difference of −17 between population and sample values. This information is presented in the first line of Table 7–3. Now suppose you were to create a small empirical sampling distribution of the mean by taking five independent samples, each of size N, calculating the mean of each sample, and then computing the mean of those five means (that is, determining the mean of this empirical sampling distribution of the mean). Suppose that mean is 96, which represents a difference of +6 between the population value and the sample estimate. This information is presented in the second line of Table 7–3. Now suppose empirical sampling distributions were created that were composed of 10, 50, 100, 1000, and then an infinite number of samples, each of size N. Notice in Table 7–3 that as the number of samples each of size N increases, the difference between the sample estimate and the population value tends to become smaller and smaller. Therefore, as the number of samples increases, the value of the sample estimate gets closer and closer to the population value until in the end, with an infinite number of samples, the estimator equals the value of the population parameter. If it does, as it does in this case, the sample statistic is defined to be an unbiased estimator of this population parameter. The sample mean, then, is an unbiased estimator of the population mean, although this fact is actually demonstrated by expectation theory, not by empirical sampling distributions, which are given here for purposes of illustration only.

It might seem that all sample statistics would be unbiased estimators of their corresponding parameters, but this is not the case. Recall that if the sample variance is defined with N in the denominator, it would not be an unbiased estimator of the population variance. In fact, through expectation theory, it can be shown that such a sample variance tends to underestimate the population variance. But if the denominator of the sample variance is composed of $N - 1$ rather than N, as it is throughout this text, then this sample value is an unbiased estimator of the population variance. Ironically, even though s^2 is an unbiased estimator of σ^2, its square root, the sample standard deviation, s, is not an unbiased estimator of the population standard deviation. However, the bias is quite small, especially if large samples are involved, and corrections are rarely used.

Consistency Another criterion of a good estimator is consistency.

> **A consistent estimator tends to get closer to the value of the population parameter as the size of the sample increases.**

The mean and variance of a sample are likely to be closer to their corresponding population values for larger than for smaller samples. Therefore, the sample mean and variance are consistent estimators of the population mean and variance, respectively.

But are not all estimators consistent—the larger the sample the closer the sample value is likely to be to the population value? The answer is no. Suppose the first score in a sample is used to estimate the population mean. It, like the sample mean, is an unbiased estimator. But the value of the first score does not tend to converge on the value of the population mean as the sample size increases, because all the remaining cases sampled after the first are irrelevant to its value.

Relative Efficiency A third criterion for a good estimator is relative efficiency.

> **A relatively efficient estimator is one whose sampling distribution has a smaller standard error than another estimator for samples of any particular size.**

For example, the sample mean and median are both unbiased estimators of the population mean in normal distributions (but not necessarily in other types of distributions). But the variability of sample means (that is, the standard error of the mean) is less than the variability of sample medians (that is, than the standard error of the median). Therefore, the mean is a relatively more efficient estimator of the population mean in normal distributions than is the sample median. Because the normal distribution is relatively common and because many more advanced statistical procedures require normal distributions, the mean is often preferred over the median because of its relative efficiency. Note, however, that this may not be the case for certain distributions that are not normal.

Sufficiency A final criterion is sufficiency.

> **A sufficient estimator is one that cannot be improved as an estimator by using any aspects of the sample data that are not already involved in its definition.**

The sample proportion is a sufficient estimator of the population proportion because its accuracy cannot be improved by considering any aspects of the data not already involved in its definition. The sufficiency of many other statistics, including the sample mean and variance, are more complicated.

Interval Estimation

Suppose that a psychologist in a school system wants to know the average IQ of students in a given high school. It is rather expensive to give an IQ test to each student, so a random sample of 25 students is tested. Suppose the sample mean is 109.

If the psychologist were required to estimate with one value the mean IQ of the population of all students in the high school, the estimate would be 109. After all, the sample mean is an unbiased estimator of the population mean, and if the sample was indeed random, one would feel somewhat confident that the sample mean of 109 was near the population mean. This is called **point estimation** because a single value is used as the estimator.

However, if you asked the psychologist whether the population mean was *exactly* 109, the answer certainly would be no. Well, how close is 109 to the population value? An approach to answering this question is to give a range of values such that you feel reasonably confident that the interval limited by these values includes the population mean. For example, the psychologist might say that the interval of 103–115 is likely to contain the population mean. This is called **interval estimation** because the estimator is an interval, not a single value.

The interval to be constructed is called a **confidence interval** and the values describing the boundaries of such an interval are called **confidence limits**. The degree of confidence in the proposition that the stated interval actually contains the population mean is indicated by a probability value. Of course, one would expect that a very large interval would be more likely to contain the population value than a very small one (everything else being equal). There are potentially any number of confidence intervals, each having a particular probability associated with it. The most commonly used confidence intervals are the "95% confidence interval" and the "99% confidence interval."

To understand how a 95% confidence interval for a sample mean is determined, consider the data from the above example (and see Figure 7–3).

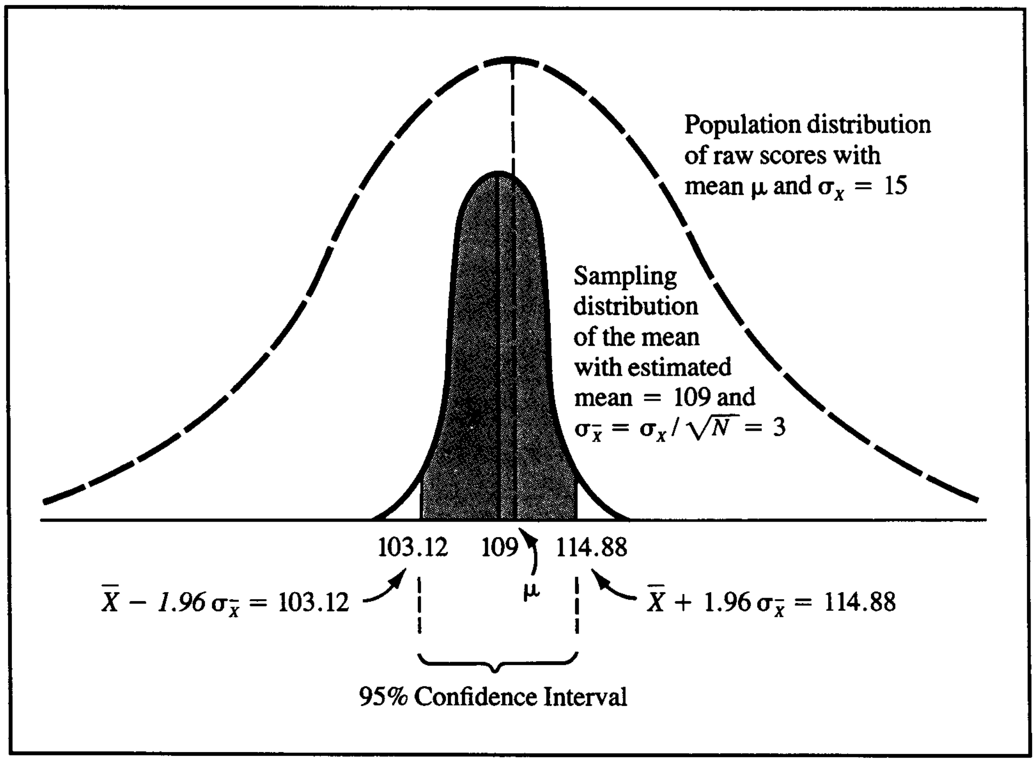

Figure 7–3 **Illustration of a 95% confidence level for a sample mean. The population of raw scores has a mean μ (dashed line) and a standard deviation of 15. A sample of 25 is drawn, with $\bar{X} = 109$. The sampling distribution of the mean is estimated to have a mean of 109 and it has a standard deviation of $\sigma_X / \sqrt{N} = 15/\sqrt{25} = 3$. The percentiles of the normal distribution indicate that 95% of such sample means would fall between $\bar{X} \pm 1.96\sigma_{\bar{X}}$, or $109 \pm 1.96(3) = 103.12$ to 114.88. This is the 95% confidence interval for the mean; it indicates that in 95% of such samples of size 25, the population mean would fall within this interval.**

The sample mean was 109 for a sample of $N = 25$. Suppose it is known that the population standard deviation is 15. Now consider the sampling distribution of the mean. Given the data at hand, the sample mean of 109 is a good estimate of the mean of the sampling distribution of the mean (that is, the population mean), and $\sigma_x / \sqrt{N} = 15/\sqrt{25} = 3$ is the standard deviation of this distribution of means (that is, the standard error of the mean, $\sigma_{\bar{x}}$). Now, 95% of the sample means will fall between $P_{.025}$ and $P_{.975}$ of the sampling distribution of the mean. The standard normal or z distribution presented in Table A of Appendix 2 can be used to determine these percentile points. To find $P_{.975}$, look down the third column in each set of three columns in Table A. This column gives the proportion of area under the standard normal to the right of a given z value or the proportion of area "in the tail" of the distribution. $P_{.975}$ is the point such that $2\frac{1}{2}\%$ or .0250 of the area falls to its right. Look down the column until you find .0250. It corresponds to the z value of 1.96. Since the standard normal distribution is symmetrical, $z = -1.96$ corresponds to $P_{.025}$. Therefore, 95% of the area of the standard normal distribution falls between the z values of -1.96 and $+1.96$. Recall, now, that the mean of the standard normal is 0 and the standard deviation is 1.00. As a result, a z-value of a point corresponds to the number of standard deviations that point is away from the mean. Therefore, 95% of the area of a normal distribution—*any* normal distribution—falls between -1.96 standard deviations and $+1.96$ standard deviations of the mean. In the present case, the normal distribution is the sampling distribution of the mean. It has a mean estimated to be $\bar{X}$ and a population standard deviation (that is, a standard error of the mean) of $\sigma_{\bar{X}}$. So, formally stated,

> **95% confidence limits for the mean of a normal distribution are $\bar{X} - 1.96\sigma_{\bar{x}}$ and $\bar{X} + 1.96\sigma_{\bar{x}}$. In words, the 95% confidence limits for the mean of a normal distribution equal the sample mean plus and minus 1.96 times the standard error of the mean.**

In this example, $\bar{X} = 109$ and $\sigma_{\bar{x}} = \sigma/\sqrt{N} = 15/\sqrt{25} = 15/5 = 3$. So the 95% confidence limits are

$$\bar{X} - 1.96\sigma_{\bar{x}} = 109 - 1.96(3) = 109 - 5.88 = 103.12$$

$$\bar{X} + 1.96\sigma_{\bar{x}} = 109 + 1.96(3) = 109 + 5.88 = 114.88.$$

Thus the probability is .95 that the interval from 103.12 to 114.88 contains the population mean. In other words, if a 95% confidence interval were computed on each of the unlimited number of samples drawn from this population, on the average 95% of such intervals would include the population mean value within their limits.

Note that the probability statement applies to the interval and not to the population mean. The population mean is a fixed value, whereas the sample mean and the confidence interval are different from sample to sample. Therefore, the statement "the probability is p that the population mean falls within the interval" is technically bad form, because it implies that the value of μ varies and might or might not happen to land in the stated interval. Actually, it is the interval which is variable, and thus a more correct statement is "the probability is p that the interval includes the population value."

Suppose one wants to be especially cautious and construct an interval that 99 times out of 100 would include the population mean. Following the same logic as above, one would go to Table A to find the z values corresponding to $P_{.995}$ and $P_{.005}$. These points are such that $\frac{1}{2}$% or .0050 of the area of the standard normal falls beyond them. Again, looking down the third column in Table A for .0050 we find .0051 and .0049, so .0050 falls halfway between. It corresponds to a z value of 2.575. Thus, 99% confidence limits for the mean are 2.575 standard errors above and below the mean. Formally,

> **99% confidence limits for the mean of a normal distribution are $\bar{X} - 2.575\sigma_{\bar{x}}$ and $\bar{X} + 2.575\sigma_{\bar{x}}$. In words, the 99% confidence limits for the mean of a normal distribution equal the sample mean plus and minus 2.575 times the standard error of the mean.**

In the present example, these limits are

$$\bar{X} - 2.575\sigma_{\bar{x}} = 109 - 2.575(3) = 109 - 7.725 = 101.275$$

$$\bar{X} + 2.575\sigma_{\bar{x}} = 109 + 2.575(3) = 109 + 7.725 = 116.725$$

Therefore, the probability is .99 that the interval 101.275 to 116.725 contains the population mean.

Two things should be observed. First, recall that the sampling distribution of the mean will be normal in form if the distribution of raw scores is normal or if the sample size is large even if the distribution of raw scores is not normal. So, in most cases, the assumption of a normal distribution will be met. Second, the population standard error is required. When it is not available and it must be estimated from sample data, the z distribution is no longer appropriate and another theoretical distribution, the Student's t distribution, is needed. This will be described in Chapter 8.

SUMMARY

Statistical inference consists of using probability to make decisions about a population on the basis of a sample of observations. A sample may be obtained by simple random sampling in which each element of the population has an equal probability of being selected for the sample. Usually the likelihood that any element is selected is independent of the likelihood that any other element will be selected. Since a sample is a subset of a population, a statistic calculated on a sample will not necessarily be the same value from sample to sample. This is sampling error. The distribution of a statistic determined on separate independent samples of size N drawn from a given population is called a sampling distribution, and the standard deviation of this distribution, called the standard error of that statistic, is a numerical index of this sampling error. Given random sampling, the sampling distribution of the mean, for example, is normal in form if the population of raw scores is normal, or it approaches normality as the size of the sample increases. This implies that the standard normal distribution can be used to determine the probability that the sample mean takes on certain values.

FORMULAS

1. **Standard score form**

$$z = \frac{X - \mu}{\sigma_x} \qquad z = \frac{\overline{X} - \mu}{\sigma_{\overline{X}}}$$

2. **Standard error of the mean**

$$\sigma_{\overline{X}} = \frac{\sigma_x}{\sqrt{N}} \text{ (population or theoretical)}$$

$$s_{\overline{X}} = \frac{s_x}{\sqrt{N}} \text{ (sample or empirical)}$$

3. **Confidence limits for the mean (population $\sigma_{\overline{X}}$ known)**

95% limits:

$$\overline{X} - 1.96\sigma_{\overline{X}} \text{ and } \overline{X} + 1.96\sigma_{\overline{X}}$$

99% limits:

$$\overline{X} - 2.575\sigma_{\overline{X}} \text{ and } \overline{X} + 2.575\sigma_{\overline{X}}$$

EXERCISES

1. Discuss the appropriateness of the following samples for the stated populations:
   a. A researcher in education wants to test the effectiveness of two different teaching methods on college students in a university, using one method with an 8 A.M. class and the other with a 2 P.M. class.
   b. A social psychologist is investigating patterns of group dynamics in the context of a jury-simulation experiment to identify the factors involved in the change of attitudes that occurs during jury debates among young adults. Students taking Introduction to Psychology may volunteer to participate in an experiment. Eighty students select the jury-simulation experiment from the 20 possible research projects available to them.
   c. An advertiser wants to know how the public is responding to the new Wonderease Soap Powder package. The package has just been displayed on television, so the advertiser quickly sets up a telephone survey in a major city to inquire whether the people who had seen the commercial liked the new package.
2. A population of 10 scores is listed below:

$$X_i = 2, 3, 3, 4, 5, 5, 7, 8, 8, 9$$

   Write each score on a piece of paper and place it in a hat. Randomly select 10 samples each of size 4 from this population, recording the values and computing the mean for each sample. (Replace the selected numbers after each drawing of 4.) Create an *empirical* sampling distribution of the mean from these data by calculating the mean of the means and the standard deviation of the means. Then estimate the *theoretical* standard error of the mean by using the first sample you collect and the appropriate formula. What does the standard error of the mean tell you about the sample mean as an estimator of the population mean?
3. For the population standard deviations given below, what is the standard error of the sampling distribution of the mean for samples of size N?
   a. If $\sigma_x = 16$, $N = 16$
   b. If $\sigma_x = 16$, $N = 64$
   c. If $\sigma_x = 50$, $N = 25$
   d. If $\sigma_x = 50$, $N = 100$
   e. If $\sigma_x = 35$, $N = 49$
   f. If $\sigma_x = 1.26$, $N = 36$
   g. If $N = 25$ and $\sigma_{\bar{X}} = 10$, what is σ_x?
4. The sampling distribution of the mean is normal in form if either one of two conditions is met. What are those two conditions?
5. Under what circumstances are the mean and variance independent?
6. What is the relationship between an idealized experiment, probability, and theoretical relative frequency?
7. In a normal distribution with $\mu = 55$ and $\sigma_x = 10$, what is the probability of randomly sampling a subject who scores
   a. 62 or higher?
   b. 70 or higher?
   c. 40 or higher? 40 or lower?
   d. between 45 and 58?
8. Suppose IQ is distributed normally in the population, with a mean of 100 and a standard deviation of 16. What is the probability of randomly sampling a person with an IQ that is
   a. 100 or higher?
   b. 100 or lower?
   c. between 84 and 116?
   d. higher than 120?
   e. between 92 and 124?
9. Assuming a normal population distribution with $\mu = 85$, $\sigma_x = 20$, what is the probability of obtaining the following means for groups of subjects chosen by random sampling?
   a. $\bar{X} = 96$ or higher, $N = 4$?
   b. $\bar{X} = 92$ or higher, $N = 16$?
   c. $\bar{X} = 78$ or higher, $N = 25$? $\bar{X} = 78$ or lower, $N = 25$?
   d. $\bar{X}$ between 74 and 83, $N = 9$?
10. Suppose it is known from previous studies that high school seniors in the country average a score of 140 with a standard deviation

of 20 on a national test of basic skills. If a counselor gives the test to all seniors in her school ($N = 169$), what is the probability that they should obtain a mean

a. higher than 142?
b. between 140 and 143?
c. lower than 139?
d. between 138 and 141?
e. between 141 and 144?

11. Define four characteristics of a good estimator.

12. Suppose we assume that scores on a national achievement test are normally distributed. If a counselor gives the test to all children in the school, determine 95% and 99% confidence intervals for the mean for the following grades:

a. Third grade: $\overline{X} = 70$, $\sigma_{\overline{X}} = 9$
b. Sixth grade: $\overline{X} = 110$, $\sigma_{\overline{X}} = 12$
c. Tenth grade: $\overline{X} = 180$, $\sigma_{\overline{X}} = 27$