CHAPTER 8

Populations, Samples, and Probability

POPULATIONS AND SAMPLES

PROBABILITY

Preview

In everyday life, we regularly generalize from limited sets of observations. One sip indicates that the batch of soup is too salty; dipping a toe in the swimming pool reassures us before taking the first plunge; a test drive triggers suspicions that the used car is not what it was advertised to be; and a casual encounter with a stranger stimulates fantasies about a deeper relationship. Valid generalizations in inferential statistics require either random sampling in the case of surveys or random assignment in the case of experiments. Introduced in this chapter, tables of random numbers can be used as aids to random sampling or random assignment.

Conclusions that we'll encounter in inferential statistics, such as "95 percent confident" or "significant at the .05 level," are statements based on probabilities. We'll define probability for a simple event and then discuss two rules for finding probabilities of more complex outcomes, including (in Review Question 8.18 on page 165) the probability of the catastrophic failure of the Challenger shuttle in 1986, which took the lives of seven astronauts.

POPULATIONS AND SAMPLES

Generalizations can backfire if a sample misrepresents the population. Faced with the possibility of erroneous generalizations, you might prefer to bypass the uncertainties of inferential statistics by surveying an entire population. This is often done if the size of the population is small. For instance, you calculate your GPA from all of your course grades, not just from a sample. If the size of the population is large, however, complete surveys are often prohibitively expensive and sometimes impossible. Under these circumstances, you might have to use samples and risk the possibility of erroneous generalizations. For instance, you might have to use a sample to estimate the mean annual income for parents of all students at a large university.

8.1 POPULATIONS

Population
Any complete set of observations (or potential observations).

Any complete set of observations (or potential observations) may be characterized as a **population**. Accurate descriptions of populations specify the nature of the observations to be taken. For example, a population might be described as "attitudes toward abortion of currently enrolled students at Bucknell University" or as "SAT critical reading scores of currently enrolled students at Rutgers University."

Real Populations

Pollsters, such as the Gallup Organization, deal with real populations. A *real* population is one in which all potential observations are accessible at the time of sampling. Examples of real populations include the two described in the previous paragraph, as well as the ages of all visitors to Disneyland on a given day, the ethnic backgrounds of all current employees of the U.S. Postal Department, and presidential preferences of all currently registered voters in the United States. Incidentally, federal law requires that a complete survey be taken every 10 years of the real population of all U.S. households—at considerable expense, involving thousands of data collectors—as a means of revising election districts for the House of Representatives. (An estimated undercount of millions of people, particularly minorities, in both the 2000 and 2010 censuses has revived a suggestion, long endorsed by statisticians, that the entire U.S. population could be estimated more accurately if a highly trained group of data collectors focused only on a random sample of households.)

WWW

INTERNET SITE
Go to the website for this book (http://www.wiley.com\college\witte). Click on the *Student Companion Site*, then *Internet Sites*, and finally ***U.S. Census Bureau*** to view its website, including links to its many reports and to population clocks that show current population estimates for the United States and the world.

Hypothetical Populations

Insofar as research workers concern themselves with populations, they often invoke the notion of a hypothetical population. A *hypothetical* population is one in which all potential observations are not accessible at the time of sampling. In most experiments, subjects are selected from very small, uninspiring real populations: the lab rats housed in the local animal colony or student volunteers from general psychology classes. Experimental subjects often are viewed, nevertheless, as a sample from a much larger hypothetical population, loosely described as "the scores of all similar animal subjects (or student volunteers) who could conceivably undergo the present experiment."

According to the rules of inferential statistics, generalizations should be made only to real populations that, in fact, have been sampled. Generalizations to hypothetical populations should be viewed, therefore, as provisional conclusions based on the wisdom of the researcher rather than on any logical or statistical necessity. In effect, it's an open question—often answered only by additional experimentation—whether or not a given experimental finding merits the generality assigned to it by the researcher.

8.2 SAMPLES

Sample
Any subset of observations from a population.

Any subset of observations from a population may be characterized as a **sample**. In typical applications of inferential statistics, the sample size is small relative to the population size. For example, less than 1 percent of all U.S. worksites are included in the Bureau of Labor Statistics' monthly survey to estimate the rate of unemployment. And although, only 1475 likely voters had been sampled in the final poll for the 2012 presidential election by the NBC News/*Wall Street Journal*, it correctly predicted that Obama would be the slim winner of the popular vote (http://www.wsj.com/election 2012).

Optimal Sample Size

There is no simple rule of thumb for determining the best or optimal sample size for any particular situation. Often sample sizes are in the hundreds or even the thousands for surveys, but they are less than 100 for most experiments. Optimal sample size depends on the answers to a number of questions, including "What is the estimated variability among observations?" and "What is an acceptable amount of error in our conclusion?" Once these types of questions have been answered, with the aid of guidelines such as those discussed in Section 11.11, specific procedures can be followed to determine the optimal sample size for any situation.

Progress Check *8.1 For each of the following pairs, indicate with a Yes or No whether the relationship between the first and second expressions could describe that between a sample and its population, respectively.

(a) students in the last row; students in class

(b) citizens of Wyoming; citizens of New York

(c) 20 lab rats in an experiment; all lab rats, similar to those used, that could undergo the same experiment

(d) all U.S. presidents; all registered Republicans

(e) two tosses of a coin; all possible tosses of a coin

Progress Check *8.2 Identify all of the expressions from Progress Check 8.1 that involve a hypothetical population.

Answers on page 429.

Random Sampling

A selection process that guarantees all potential observations in the population have an equal chance of being selected.

8.3 RANDOM SAMPLING

The valid use of techniques from inferential statistics requires that samples be random.

> ***Random sampling*** **occurs if, at each stage of sampling, the selection process guarantees that all potential observations in the population have an equal chance of being included in the sample.**

It's important to note that randomness describes the *selection process*—that is, the conditions under which the sample is taken—and not the particular pattern of observations in the sample. Having established that sampling is random, you still can't predict anything about the unique pattern of observations in that sample. The observations in the sample should be representative of those in the population, but there is no guarantee that they actually will be.

Casual or Haphazard, Not Random

A casual or haphazard sample doesn't qualify as a random sample. Not every student at UC San Diego has an equal chance of being sampled if, for instance, a pollster casually selects only students who enter the student union. Obviously excluded from this sample are all those students (few, we hope) who never enter the student union. Even the final selection of students from among those who do enter the student union might reflect the pollster's various biases, such as an unconscious preference for attractive students who are walking alone.

Progress Check * 8.3 Indicate whether each of the following statements is True or False. A random selection of 10 playing cards from a deck of 52 cards implies that

(a) the random sample of 10 cards accurately represents the important features of the whole deck.

(b) each card in the deck has an equal chance of being selected.

(c) it is impossible to get 10 cards from the same suit (for example, 10 hearts).

(d) any outcome, however unlikely, is possible.

Answers on page 429.

8.4 TABLES OF RANDOM NUMBERS

Tables of random numbers can be used to obtain a random sample. These tables are generated by a computer designed to equalize the occurrence of any one of the 10 digits: 0, 1, 2, . . . , 8, 9. For convenience, many random number tables are spaced in columns of five-digit numbers. Table H in Appendix C shows a specimen page of random numbers from a book devoted entirely to random digits.

How Many Digits?

The size of the population determines whether you deal with numbers having one, two, three, or more digits. The only requirement is that you have at least as many different numbers as you have potential observations within the population. For example, if you were attempting to take a random sample from a population consisting of 679 students at some college, you could use the 1000 three-digit numbers ranging

8.10 PROBABILITY AND STATISTICS

Probability assumes a key role in inferential statistics including, for instance, the important area known as *hypothesis testing*. Because of the inevitable variability that accompanies any observed result, such as a mean difference between two groups, its value must be viewed within the context of the many possible results that could have occurred just by chance. With the aid of some theoretical curve, such as the normal curve, and a provisional assumption, known as the *null hypothesis,* that chance can reasonably account for the result, probabilities are assigned to the one observed mean difference. If this probability is very small, the result is viewed as a rare outcome, and we conclude that something real—that is, something that can't reasonably be attributed to chance—has occurred. On the other hand, if this probability isn't very small, the result is viewed as a common outcome, and we conclude that something transitory—that is, something that can reasonably be attributed to chance—has occurred.

Common Outcomes

Common outcomes signify, most generally, a lack of evidence that something special has occurred. For instance, they suggest that the observed mean difference—whatever its value—might signify that the true mean difference could equal zero and, therefore, that any comparable study would just as likely produce either a positive or negative mean difference. Therefore, the observed mean difference should not be taken seriously because, in the language of statistics, it lacks *statistical significance.*

Rare Outcomes

On the other hand, rare outcomes signify that something special has occurred. For instance, they suggest that the observed mean difference probably signifies a true mean difference equal to some nonzero value and, therefore, that any comparable study would most likely produce a mean difference with the same sign and a value in the neighborhood of the one originally observed. Therefore, the observed mean difference should be taken seriously because it has statistical significance.

Common or Rare?

As an aid to determining whether observed results should be viewed as common or rare, statisticians interpret different *proportions of area under theoretical curves,* such as the normal curve shown in **Figure 8.2**, *as probabilities of random outcomes.* For instance, the standard normal table indicates that .9500 is the proportion of total area between z scores of –1.96 and +1.96. (Verify this proportion by referring to Table A in Appendix C and, if necessary, to the latter part of Section 5.6.) Accordingly, the probability of a randomly selected z score anywhere between ±1.96 equals .95. Because it should happen about 95 times out of 100, this is often designated as a *common* event signifying that, once variability is considered, nothing special is happening. On the other hand, since the standard normal curve indicates that .025 is the proportion of total area above a z score of +1.96, and also that .025 is the proportion of total area below a z score of –1.96, then the probability of a randomly selected z score anywhere beyond either +1.96 or –1.96 equals .05 (from .025 + .025, thanks to the addition rule). Because it should happen only about 5 times in 100, this is often designated as a *rare* outcome signifying that something special is happening.

At this point, you're not expected to understand the rationale behind this perspective, but merely that, once identified with a particular result, a specified sector of area under a curve will be interpreted as the probability of that outcome. Furthermore, since the probability of an outcome has important implications for generalizing beyond actual results, probabilities play a key role in inferential statistics.

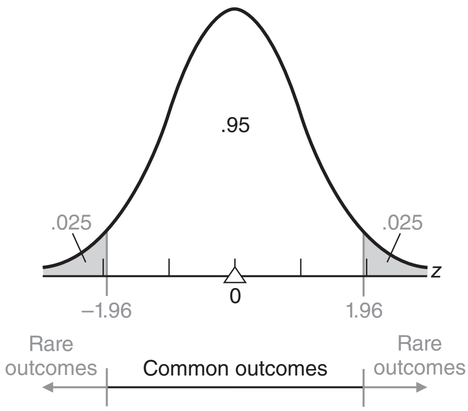

FIGURE 8.2
One possible model for determining common and rare outcomes.

Progress Check *8.9 Referring to the standard normal table (Table A, Appendix C), find the probability that a randomly selected *z* score will be

(a) above 1.96

(b) either above 1.96 or below –1.96

(c) between –1.96 and 1.96

(d) either above 2.58 or below –2.58

Answers on page 431.

Summary

Any set of potential observations may be characterized as a population. Any subset of observations constitutes a sample.

Populations are either real or hypothetical, depending on whether or not all observations are accessible at the time of sampling.

The valid application of techniques from inferential statistics requires that the samples be random or that subjects be randomly assigned. A sample is random if at each stage of sampling the selection process guarantees that all remaining observations in the population have an equal chance of being included in the sample. Random assignment occurs whenever all subjects have an equal opportunity of being assigned to each of the various groups.

Tables of random numbers provide one method both for taking random samples in surveys and for randomly assigning subjects to various groups in experiments. Some type of randomization always should occur during the early stages of any investigation, whether a survey or an experiment.

The probability of an event specifies the proportion of times that this event is likely to occur.

Whenever you must find the probability of sets of mutually exclusive events connected with the word *or,* use the addition rule: Add together the separate probabilities of each of the mutually exclusive events to find the probability that any one of these events will occur. Whenever events aren't mutually exclusive, the addition rule must be adjusted for the overlap between outcomes.

Whenever you must find the probability of sets of independent events connected with the word *and,* use the multiplication rule: Multiply together the separate probabilities of each of the independent events to find the probability that these events will occur together. Whenever events are dependent, the multiplication rule must be adjusted by using the conditional probability of the second outcome, given the occurrence of the first outcome.

In inferential statistics, sectors of area under various theoretical curves are interpreted as probabilities, and these probabilities play a key role in inferential statistics.

Important terms

Population	**Sample**
Random sampling	**Random assignment**
Probability	**Mutually exclusive events**
Addition rule	**Independent events**
Multiplication rule	**Conditional probability**

Key equations

ADDITION RULE

$$\Pr(A \text{ or } B) = \Pr(A) + \Pr(B)$$

MULTIPLICATION RULE

$$\Pr(A \text{ and } B) = [\Pr(A)][\Pr(B)]$$

REVIEW QUESTIONS

8.10 Television stations sometimes solicit feedback volunteered by viewers about a televised event. Following a televised debate between Barack Obama and Mitt Romney in the 2012 presidential election campaign, a TV station conducted a telephone poll to determine the “winner.” Callers were given two phone numbers, one for Obama and the other for Romney, to register their opinions automatically.

(a) Comment on whether or not this was a random sample.

(b) How might this poll have been improved?

8.11 You want to take a random sample of 30 from a population described by telephone directory with a single telephone area code. Indicate whether or not each of the following selection techniques would be a random sample and, if not, why. Using the telephone directory,

(a) make 30 blind pencil stabs.

(b) refer to tables of random numbers to determine the page and then the position of the selected person on that page. Repeat 30 times.

(c) refer to tables of random numbers to find six-digit numbers that identify the page number and line on that page for each of 30 people.

(d) select 30 people haphazardly.

CHAPTER 9

Sampling Distribution of the Mean

Preview

This chapter focuses on the single most important concept in inferential statistics—the concept of a sampling distribution. A sampling distribution serves as a frame of reference for every outcome, among all possible outcomes, that could occur just by chance. It reappears in every subsequent chapter as the key to understanding how, once variability has been estimated, we can generalize beyond a limited set of actual observations. In order to use a sampling distribution, we must identify its mean, its standard deviation, and its shape—a seemingly difficult task that, thanks to the theory of statistics, can be performed by invoking the population mean, the population standard deviation, and the normal curve, respectively.

There's a good chance that you've taken the SAT test, and you probably remember your scores. Assume that the SAT math scores for all college-bound students during a recent year were distributed around a mean of 500 with a standard deviation of 110. An investigator at a university wishes to test the claim that, on the average, the SAT math scores for local freshmen equals the national average of 500. His task would be straightforward if, in fact, the math scores for all local freshmen were readily available. Then, after calculating the mean score for all local freshmen, a direct comparison would indicate whether, on the average, local freshmen score below, at, or above the national average.

Assume that it is not possible to obtain scores for the entire freshman class. Instead, SAT math scores are obtained for a random sample of 100 students from the local population of freshmen, and the mean score for this sample equals 533. If each sample were an exact replica of the population, generalizations from the sample to the population would be most straightforward. Having observed a mean score of 533 for a sample of 100 freshmen, we could have concluded, without even a pause, that the mean math score for the entire freshman class also equals 533 and, therefore, exceeds the national average.

9.1 WHAT IS A SAMPLING DISTRIBUTION?

Random samples rarely represent the underlying population exactly. Even a mean math score of 533 could originate, just by chance, from a population of freshmen whose mean equals the national average of 500. Accordingly, generalizations from a single sample to a population are much more tentative. Indeed, generalizations are based not merely on the single sample mean of 533 but also on its distribution—a distribution of sample means for all possible random samples. Representing the statistician's model of random outcomes,

the *sampling distribution of the mean* refers to the probability distribution of means for all possible random samples of a given size from some population.

Sampling Distribution of the Mean
Probability distribution of means for all possible random samples of a given size from some population.

In effect, this distribution describes the variability among sample means that could occur just by chance and thereby serves as a frame of reference for generalizing from a single sample mean to a population mean.

The sampling distribution of the mean allows us to determine whether, given the variability among all possible sample means, the one observed sample mean can be viewed as a *common* outcome or as a *rare* outcome (from a distribution centered, in this case, about a value of 500). If the sample mean of 533 qualifies as a *common* outcome in this sampling distribution, then the difference between 533 and 500 isn't large enough, relative to the variability of all possible sample means, to signify that anything special is happening in the underlying population. Therefore, we can conclude that the mean math score for the entire freshman class could be the same as the national average of 500. On the other hand, if the sample mean of 533 qualifies as a *rare* outcome in this sampling distribution, then the difference between 533 and 500 is large enough, relative to the variability of all possible sample means, to signify that something special probably is happening in the underlying population. Therefore, we can conclude that the mean math score for the entire freshman class probably exceeds the national average of 500.

All Possible Random Samples

When attempting to generalize from a single sample mean to a population mean, we must consult the sampling distribution of the mean. In the present case, this distribution is based on *all possible* random samples, each of size 100 that can be taken from the local population of freshmen. *All possible random samples* refers not to the number of samples of size 100 required to *survey completely* the local population of freshmen but to the number of different ways in which a *single* sample of size 100 can be selected from this population.

"All possible random samples" tends to be a huge number. For instance, if the local population contained at least 1,000 freshmen, the total number of possible random samples, each of size 100, would be astronomical in size. The 301 digits in this number would dwarf even the national debt. Even with the aid of a computer, it would be a horrendous task to construct this sampling distribution from scratch, itemizing each mean for all possible random samples.

Fortunately, statistical theory supplies us with considerable information about the sampling distribution of the mean, as will be discussed in the remainder of this chapter. Armed with this information about sampling distributions, we'll return to the current example in the next chapter and test the claim that the mean math score for the local population of freshmen equals the national average of 500. Only at that point—and not at the end of this chapter—should you expect to understand completely the role of sampling distributions in practical applications.

9.2 CREATING A SAMPLING DISTRIBUTION FROM SCRATCH

Let's establish precisely what constitutes a sampling distribution by creating one from scratch under highly simplified conditions. Imagine some ridiculously small population of four observations with values of 2, 3, 4, and 5, as shown in **Figure 9.1.** Next, itemize all possible random samples, each of size two, that could be taken from this population. There are four possibilities on the first draw from the population and also four possibilities on the second draw from the population, as indicated in **Table 9.1.*** The two sets of possibilities combine to yield a total of 16 possible samples. At this point, remember, we're clarifying the notion of a sampling distribution of the mean. In practice, only a single random sample, not 16 possible samples, would be taken from the population; the sample size would be very small relative to a much larger population size, and, of course, not all observations in the population would be known.

For each of the 16 possible samples, Table 9.1 also lists a sample mean (found by adding the two observations and dividing by 2) and its probability of occurrence (expressed as $\frac{1}{16}$, since each of the 16 possible samples is equally likely). When cast into a relative frequency or probability distribution, as in **Table 9.2,** the 16 sample means constitute the sampling distribution of the mean, previously defined as the probability distribution of means for all possible random samples of a given size from some population. Not all values of the sample mean occur with equal probabilities in Table 9.2 since some values occur more than once among the 16 possible samples. For instance, a sample mean value of 3.5 appears among 4 of 16 possibilities and has a probability of $\frac{4}{16}$.

Probability of a Particular Sample Mean

The distribution in Table 9.2 can be consulted to determine the probability of obtaining a particular sample mean or set of sample means. For example, the probability of a randomly selected sample mean of 5.0 equals $\frac{1}{16}$ or .0625. According to the addition

*Ordinarily, a single observation is sampled only once, that is, sampling is *without replacement*. If employed with the present, highly simplified example, however, sampling without replacement would magnify an unimportant technical adjustment.

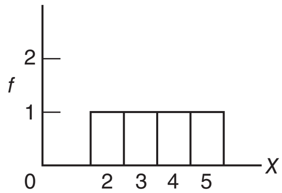

FIGURE 9.1
Graph of a miniature population.

Table 9.1
ALL POSSIBLE SAMPLES OF SIZE TWO FROM A MINIATURE POPULATION

	ALL POSSIBLE SAMPLES	MEAN ($\overline{X}$)	PROBABILITY
(1)	2,2	2.0	$\frac{1}{16}$
(2)	2,3	2.5	$\frac{1}{16}$
(3)	2,4	3.0	$\frac{1}{16}$
(4)	2,5	3.5	$\frac{1}{16}$
(5)	3,2	2.5	$\frac{1}{16}$
(6)	3,3	3.0	$\frac{1}{16}$
(7)	3,4	3.5	$\frac{1}{16}$
(8)	3,5	4.0	$\frac{1}{16}$
(9)	4,2	3.0	$\frac{1}{16}$
(10)	4,3	3.5	$\frac{1}{16}$
(11)	4,4	4.0	$\frac{1}{16}$
(12)	4,5	4.5	$\frac{1}{16}$
(13)	5,2	3.5	$\frac{1}{16}$
(14)	5,3	4.0	$\frac{1}{16}$
(15)	5,4	4.5	$\frac{1}{16}$
(16)	5,5	5.0	$\frac{1}{16}$

rule for mutually exclusive outcomes, described in Chapter 8, the probability of a randomly selected sample mean of either 5.0 or 2.0 equals $\frac{1}{16} + \frac{1}{16} = \frac{2}{16} = .1250$. This type of probability statement, based on a sampling distribution, assumes an essential role in inferential statistics and will reappear throughout the remainder of the book.

Review

Figure 9.2 summarizes the previous discussion. It depicts the emergence of the sampling distribution of the mean from the set of all possible (16) samples of size two,

Table 9.2
SAMPLING DISTRIBUTION OF THE MEAN (SAMPLES OF SIZE TWO FROM A MINIATURE POPULATION)

SAMPLE MEAN ($\overline{X}$)	PROBABILITY
5.0	$\frac{1}{16}$
4.5	$\frac{2}{16}$
4.0	$\frac{3}{16}$
3.5	$\frac{4}{16}$
3.0	$\frac{3}{16}$
2.5	$\frac{2}{16}$
2.0	$\frac{1}{16}$

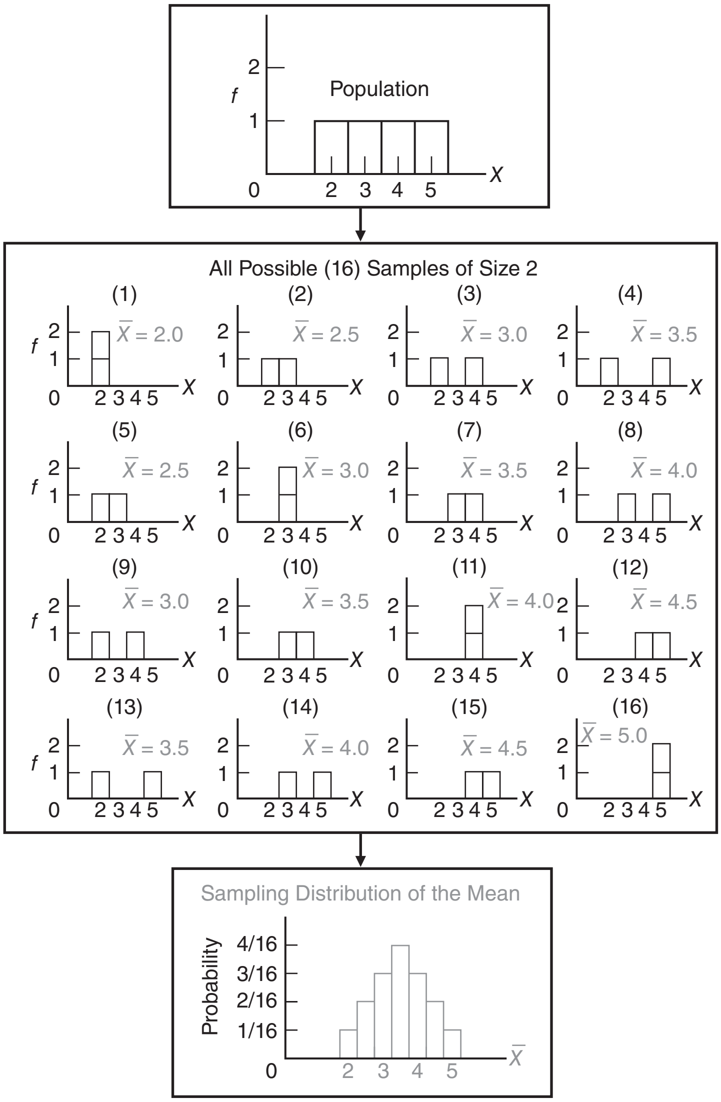

FIGURE 9.2
Emergence of the sampling distribution of the mean from all possible samples.

based on the miniature population of four observations. Familiarize yourself with this figure, as it will be referred to again.

Progress Check *9.1 Imagine a very simple population consisting of only five observations: 2, 4, 6, 8, 10.

(a) List all possible samples of size two.

(b) Construct a relative frequency table showing the sampling distribution of the mean.

Answers on pages 431 and 432.

Reminder:
We recommend memorizing the symbols in Table 9.3.

Table 9.3
SYMBOLS FOR THE MEAN AND STANDARD DEVIATION OF THREE TYPES OF DISTRIBUTIONS

TYPE OF DISTRIBUTION	MEAN	STANDARD DEVIATION
Sample	$\overline{X}$	s
Population	μ	σ
Sampling distribution of the mean	$\mu_{\overline{X}}$	$\sigma_{\overline{X}}$ (standard error of the mean)

9.3 SOME IMPORTANT SYMBOLS

Having established precisely what constitutes a sampling distribution under highly simplified conditions, we can introduce the special symbols that identify the mean and the standard deviation of the sampling distribution of the mean. **Table 9.3** also lists the corresponding symbols for the sample and the population. It would be wise to memorize these symbols.

You are already acquainted with the English letters $\overline{X}$ and s, representing the mean and standard deviation of any sample, and also the Greek letters μ (mu) and σ (sigma), representing the mean and standard deviation of any population. New are the Greek letters $\mu_{\overline{X}}$ (mu sub X-bar) and $\sigma_{\overline{X}}$ (sigma sub X-bar), representing the mean and standard deviation, respectively, of the sampling distribution of the mean. To minimize confusion, the latter term, $\sigma_{\overline{X}}$, is often referred to as the *standard error of the mean* or simply as the *standard error*.

Significance of Greek Letters

Note that Greek letters are used to describe characteristics of both populations and sampling distributions, suggesting a common feature. Both types of distribution deal with all possibilities, that is, with *all possible observations* in the population, or with the *means of all possible random samples* in the sampling distribution of the mean.

With this background, let's focus on the three most important characteristics of the sampling distribution of the mean: its mean, its standard deviation, and its shape. In subsequent chapters, these three characteristics will form the basis for applied work in inferential statistics.

Progress Check *9.2 Without peeking, list the special symbols for the mean of the population ___(a)___, mean of the sampling distribution of the mean ___(b)___, mean of the sample ___(c)___, standard error of the mean ___(d)___, standard deviation of the sample ___(e)___, and standard deviation of the population ___(f)___.

Answers on page 432.

9.4 MEAN OF ALL SAMPLE MEANS ($\mu_{\overline{X}}$)

The distribution of sample means itself has a mean.

> **The *mean of the sampling distribution of the mean* always equals the mean of the population.**

Expressed in symbols, we have

Mean of the Sampling Distribution of the Mean ($\mu_{\overline{X}}$)
The mean of all sample means always equals the population mean.

MEAN OF THE SAMPLING DISTRIBUTION

$$\mu_{\overline{X}} = \mu \qquad (9.1)$$

where $\mu_{\overline{X}}$ represents the mean of the sampling distribution and μ represents the mean of the population.

Interchangeable Means

Since the mean of all sample means ($\mu_{\overline{X}}$) always equals the mean of the population (μ), these two terms are interchangeable in inferential statistics. Any claims about the population mean can be transferred directly to the mean of the sampling distribution, and vice versa. If, as claimed, the mean math score for the local population of freshmen equals the national average of 500, then the mean of the sampling distribution also automatically will equal 500. For the same reason, it's permissible to view the one observed sample mean of 533 as a deviation either from the mean of the sampling distribution or from the mean of the population. It should be apparent, therefore, that *whether an expression involves $\mu_{\overline{X}}$ or μ, it reflects, at most, a difference in emphasis on either the sampling distribution or the population, respectively, rather than any difference in numerical value.*

Explanation

Although important, it's not particularly startling that the mean of all sample means equals the population mean. As can be seen in Figure 9.2, samples are not exact replicas of the population, and most sample means are either larger or smaller than the population mean (equal to 3.5 in Figure 9.2). By taking the mean of all sample means, however, you effectively neutralize chance differences between sample means and retain a value equal to the population mean.

Progress Check *9.3 Indicate whether the following statements are True or False. The mean of all sample means, $\mu_{\overline{X}}$, . . .

(a) always equals the value of a particular sample mean.

(b) equals 100 if, in fact, the population mean equals 100.

(c) usually equals the value of a particular sample mean.

(d) is interchangeable with the population mean.

Answers on page 432.

9.5 STANDARD ERROR OF THE MEAN ($\sigma_{\overline{X}}$)

The distribution of sample means also has a standard deviation, referred to as the standard error of the mean.

The *standard error of the mean* equals the standard deviation of the population divided by the square root of the sample size.

Standard Error of the Mean ($\sigma_{\overline{X}}$)
A rough measure of the average amount by which sample means deviate from the mean of the sampling distribution or from the population mean.

Expressed in symbols,

STANDARD ERROR OF THE MEAN

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} \tag{9.2}$$

where $\sigma_{\overline{X}}$ represents the standard error of the mean; σ represents the standard deviation of the population; and n represents the sample size.

Special Type of Standard Deviation

The standard error of the mean serves as a special type of standard deviation that measures variability in the sampling distribution. It supplies us with a *standard,* much like a yardstick, that describes the amount by which sample means deviate from the mean of the sampling distribution or from the population mean. The *error* in standard error refers not to computational errors, but to errors in generalizations attributable to the fact that, just by chance, most random samples aren't exact replicas of the population.

You might find it helpful to think of the standard error of the mean as a rough measure of the average amount by which sample means deviate from the mean of the sampling distribution or from the population mean.

Insofar as the shape of the distribution sample means approximates a normal curve, as described in the next section, about 68 percent of all sample means deviate less than one standard error from the mean of the sampling distribution, whereas only about 5 percent of all sample means deviate more than two standard errors from the mean of this distribution.

Effect of Sample Size

A most important implication of Formula 9.2 is that whenever the sample size equals two or more, the variability of the sampling distribution is less than that in the population. A modest demonstration of this effect appears in Figure 9.2, where the means of all possible samples cluster closer to the population mean (equal to 3.5) than do the four original observations in the population. A more dramatic demonstration occurs with larger sample sizes. Earlier in this chapter, for instance, 110 was given as the value of σ, the population standard deviation for SAT scores. Much smaller is the variability in the sampling distribution of mean SAT scores, each based on samples of 100 freshmen. According to Formula 9.2, in the present example,

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{110}{\sqrt{100}} = \frac{110}{10} = 11$$

there is a tenfold reduction in variability, from 110 to 11, when our focus shifts from the population to the sampling distribution.

According to Formula 9.2, any increase in sample size translates into a smaller standard error and, therefore, into a *new* sampling distribution with less variability. With a larger sample size, sample means cluster more closely about the mean of the sampling distribution and about the mean of the population and, therefore, allow more precise generalizations from samples to populations.

Explanation

It's not surprising that variability should be smaller in sampling distributions than in populations. The population standard deviation reflects variability among *individual observations,* and it is directly affected by any relatively large or small observations within the population. On the other hand, the standard error of the mean reflects variability among *sample means,* each of which represents a collection of individual observations. The appearance of relatively large or small observations within a particular sample tends to affect the sample mean only slightly, because of the stabilizing presence in the same sample of other, more moderate observations or even extreme observations in the opposite direction. This stabilizing effect becomes even more pronounced with larger sample sizes.

Progress Check *9.4 Indicate whether the following statements are True or False. The standard error of the mean, $\sigma_{\overline{X}}$, . . .

(a) roughly measures the average amount by which sample means deviate from the population mean.

(b) measures variability in a particular sample.

(c) increases in value with larger sample sizes.

(d) equals 5, given that $\sigma = 40$ and $n = 64$.

Answers on page 432.

9.6 SHAPE OF THE SAMPLING DISTRIBUTION

Central Limit Theorem

Regardless of the population shape, the shape of the sampling distribution of the mean approximates a normal curve if the sample size is sufficiently large.

A product of statistical theory, expressed in its simplest form,

> **the *central limit theorem* states that, regardless of the shape of the population, the shape of the sampling distribution of the mean approximates a normal curve *if the sample size is sufficiently large*.**

According to this theorem, it doesn't matter whether the shape of the parent population is normal, positively skewed, negatively skewed, or some nameless, bizarre shape, as long as the sample size is sufficiently large. What constitutes "sufficiently large" depends on the shape of the parent population. If the shape of the parent population is normal, then any sample size (even a sample size of one) will be sufficiently large. Otherwise, depending on the degree of non-normality in the parent population, a sample size between 25 and 100 is sufficiently large.

Examples

For the population with a non-normal shape in the top panel of Figure 9.2, the shape of the sampling distribution in the bottom panel reveals a preliminary drift toward normality—that is, a shape having a peak in the middle with tapered flanks on either side—even for very small samples of size 2. For the two non-normal populations in the top panel of **Figure 9.3**, the shapes of the sampling distributions in the middle panel show essentially the same preliminary drift toward normality when the sample size equals only 2, while the shapes of the sampling distributions in the bottom panel closely approximate normality when the sample size equals 25.

Earlier in this chapter, 533 was given as the mean SAT math score for a random sample of 100 freshmen. Because this sample size satisfies the requirements of the central limit theorem, we can view the sample mean of 533 as originating from a sampling

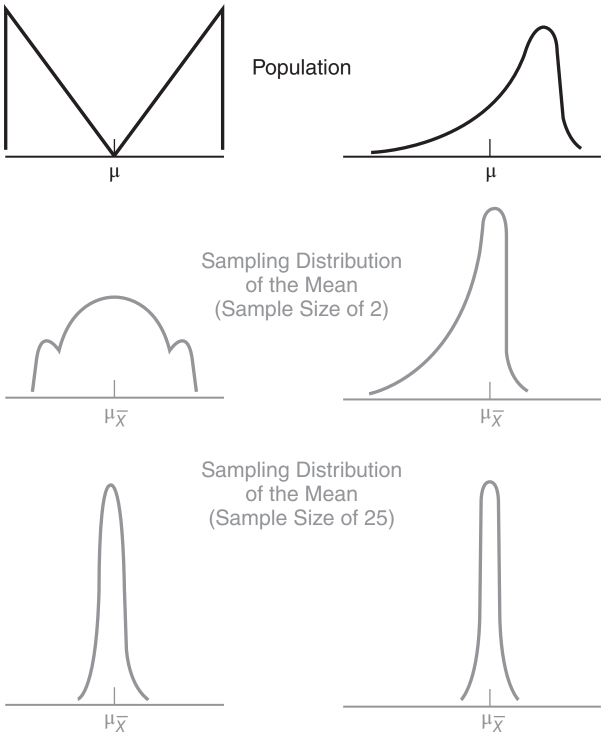

FIGURE 9.3
Effect of the central limit theorem.

distribution whose shape approximates a normal curve, even though we lack information about the shape of the population of math scores for the entire freshman class. It will be possible, therefore, to make precise statements about this sampling distribution, as described in the next chapter, by referring to the table for the standard normal curve.

Why the Central Limit Theorem Works

In a normal curve, you will recall, intermediate values are the most prevalent, and extreme values, either larger or smaller, occupy the tapered flanks. Why, when the sample size is large, does the sampling distribution approximate a normal curve, even though the parent population might be non-normal?

Many Sample Means with Intermediate Values

When the sample size is large, it is *most likely* that any single sample will contain the full spectrum of small, intermediate, and large scores from the parent population, *whatever its shape*. The calculation of a mean for this type of sample tends to neutralize or dilute the effects of any extreme scores, and the sample mean emerges with some intermediate value. Accordingly, intermediate values prevail in the sampling distribution, and they cluster around a peak frequency representing the most common or modal value of the sample mean, as suggested at the bottom of Figure 9.3.

Few Sample Means with Extreme Values

To account for the rarer sample mean values in the tails of the sampling distribution, focus on those relatively infrequent samples that, just by chance, contain less than the full spectrum of scores from the parent population. Sometimes, because of the relatively large number of extreme scores in a particular direction, the calculation of a mean only slightly dilutes their effect, and the sample mean emerges with some more extreme value. The likelihood of obtaining extreme sample mean values declines with the extremity of the value, producing the smoothly tapered, slender tails that characterize a normal curve.

Progress Check *9.5 Indicate whether the following statements are True or False. The central limit theorem

(a) states that, with sufficiently large sample sizes, the shape of the population is normal.

(b) states that, regardless of sample size, the shape of the sampling distribution of the mean is normal.

(c) ensures that the shape of the sampling distribution of the mean equals the shape of the population.

(d) applies to the shape of the sampling distribution—not to the shape of the population and not to the shape of the sample.

Answers on page 432.

9.7 OTHER SAMPLING DISTRIBUTIONS

For the Mean

There are many different sampling distributions of means. A new sampling distribution is created by a switch to another population. Furthermore, for any single population, there are as many different sampling distributions as there are possible sample sizes. Although each of these sampling distributions has the same mean, the value of the standard error always differs and depends upon the size of the sample.

For Other Measures

There are sampling distributions for measures other than a single mean. For instance, there are sampling distributions for medians, proportions, standard deviations, variances, and correlations, as well as for differences between pairs of means, pairs of proportions, and so forth. We'll have occasion to work with some of these distributions in later chapters.

Summary

The notion of a sampling distribution is the most important concept in inferential statistics. The sampling distribution of the mean is defined as the probability distribution of means for all possible random samples of a given size from some population.

Statistical theory pinpoints three important characteristics of the sampling distribution of the mean:

- The mean of the sampling distribution equals the mean of the population.
- The standard deviation of the sampling distribution, that is, the standard error of the mean, equals the standard deviation of the population divided by the square root of the sample size. An important implication of this formula is that a larger sample size translates into a sampling distribution with a smaller variability,

allowing more precise generalizations from samples to populations. The standard error of the mean serves as a rough measure of the average amount by which sample means deviate from the mean of the sampling distribution or from the population mean.

- According to the central limit theorem, regardless of the shape of the population, the shape of the sampling distribution approximates a normal curve if the sample size is sufficiently large. Depending on the degree of non-normality in the parent population, a sample size of between 25 and 100 is sufficiently large.

Any single sample mean can be viewed as originating from a sampling distribution whose (1) mean equals the population mean (whatever its value); whose (2) standard error equals the population standard deviation divided by the square root of the sample size; and whose (3) shape approximates a normal curve (if the sample size satisfies the requirements of the central limit theorem).

Important Terms

Mean of the sampling distribution of the mean ($\mu_{\overline{X}}$)
Sampling distribution of the mean
Standard error of the mean ($\sigma_{\overline{X}}$)
Central limit theorem

Key Equations

SAMPLING DISTRIBUTION MEAN

$$\mu_{\overline{X}} = \mu$$

STANDARD ERROR

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

REVIEW QUESTIONS

9.6 A random sample tends not to be an exact replica of its parent population. This fact has a number of implications. Indicate which are true and which are false.

(a) All possible random samples can include a few samples that are exact replicas of the population, but most samples aren't exact replicas.

(b) A more representative sample can be obtained by handpicking (rather than randomly selecting) observations.

(c) Insofar as it misrepresents the parent population, a random sample can cause an erroneous generalization.

(d) In practice, the mean of a single random sample is evaluated relative to the variability of means for all possible random samples.

9.7 Define the sampling distribution of the mean.

9.8 Specify three important properties of the sampling distribution of the mean.

9.9 Indicate whether the following statements are true or false. If we took a random sample of 35 subjects from some population, the associated sampling distribution of the mean would have the following properties:

(a) Shape would approximate a normal curve.

(b) Mean would equal the one sample mean.

(c) Shape would approximate the shape of the population.

(d) Compared to the population variability, the variability would be reduced by a factor equal to the square root of 35.

(e) Mean would equal the population mean.

(f) Variability would equal the population variability.

9.10 Indicate whether the following statements are true or false. The sampling distribution of the mean

(a) is always constructed from scratch, even when the population is large.

(b) serves as a bridge to aid generalizations from a sample to a population.

(c) is the same as the sample mean.

(d) always reflects the shape of the underlying population.

(e) has a mean that always coincides with the population mean.

(f) is a device used to determine the effect of variability (that is, what can happen, just by chance, when samples are random).

(g) remains unchanged even with shifts to a new population or sample size.

(h) supplies a spectrum of possibilities against which to evaluate the one observed sample mean.

(i) tends to cluster more closely about the population mean with increases in sample size.

9.11 Someone claims that, since the mean of the sampling distribution equals the population mean, any single sample mean must also equal the population mean. Any comment?

9.12 Given that population standard deviation equals 24, how large must the sample size, *n*, be in order for the standard error to equal

(a) 8 ?

(b) 6 ?

(c) 3 ?

(d) 2 ?

9.13 Given a sample size of 36, how large does the population standard deviation have to be in order for the standard error to be

(a) 1 ?

(b) 2 ?

(c) 5 ?

(d) 100 ?

9.14 **(a)** A random sample of size 144 is taken from the local population of grade-school children. Each child estimates the number of hours per week spent watching TV. At this point, what can be said about the sampling distribution?

(b) Assume that a standard deviation, σ, of 8 hours describes the TV estimates for the local population of schoolchildren. At this point, what can be said about the sampling distribution?

(c) Assume that a mean, μ, of 21 hours describes the TV estimates for the local population of schoolchildren. Now what can be said about the sampling distribution?

(d) Roughly speaking, the sample means in the sampling distribution should deviate, on average, about ___ hours from the mean of the sampling distribution and from the mean of the population.

(e) About 95 percent of the sample means in this sampling distribution should be between ___ hours and ___ hours.

CHAPTER 10

Introduction to Hypothesis Testing: The *z* Test

Preview

This chapter describes the first in a series of hypothesis tests. Learning the vocabulary of special terms for hypothesis tests will be most helpful throughout the remainder of the book. However, do not become so concerned about either terminology or computational mechanics that you lose sight of the essential role of the sampling distribution—the model of everything that could happen just by chance—in any hypothesis test.

Using the sampling distribution as our frame of reference, the one observed outcome is characterized as either a common outcome or a rare outcome. A common outcome is readily attributable to chance, and therefore, the hypothesis that nothing special is happening—the null hypothesis—is retained. On the other hand, a rare outcome isn't readily attributable to chance, and therefore, the null hypothesis is rejected (usually to the delight of the researcher).

10.1 TESTING A HYPOTHESIS ABOUT SAT SCORES

In the previous chapter, we postponed a test of the hypothesis that the mean SAT math score for all local freshmen equals the national average of 500. Now, given a mean math score of 533 for a random sample of 100 freshmen, let's test the hypothesis that, with respect to the national average, nothing special is happening in the local population. Insofar as an investigator usually suspects just the opposite—namely, that something special is happening in the local population—he or she hopes to reject the hypothesis that nothing special is happening, henceforth referred to as the *null* hypothesis and defined more formally in a later section.

Hypothesized Sampling Distribution

If the null hypothesis is true, then the distribution of sample means—that is, the sampling distribution of the mean for all possible random samples, each of size 100, from the local population of freshmen—will be centered about the national average of 500. (Remember, the mean of the sampling distribution always equals the population mean.) In **Figure 10.1**, this sampling distribution is referred to as the *hypothesized* sampling distribution, since its mean equals 500, the hypothesized mean reading score for the local population of freshmen.

Anticipating the key role of the hypothesized sampling distribution in our hypothesis test, let's focus on two more properties of this distribution:

1. In Figure 10.1, vertical lines appear, at intervals of size 11, on either side of the hypothesized population mean of 500. These intervals reflect the size of the standard error of the mean, $\sigma_{\overline{X}}$. To verify this fact, originally demonstrated in Chapter 9, substitute 110 for the population standard deviation, σ, and 100 for the sample size, n, in Formula 9.2 to obtain

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{110}{\sqrt{100}} = \frac{110}{10} = 11$$

2. Notice that the shape of the hypothesized sampling distribution in Figure 10.1 approximates a normal curve, since the sample size of 100 is large enough to satisfy the requirements of the central limit theorem. Eventually, with the aid of normal curve tables, we will be able to construct boundaries for common and rare outcomes under the null hypothesis.

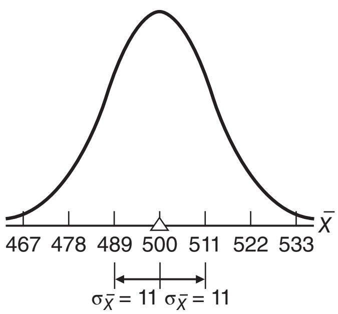

FIGURE 10.1
Hypothesized sampling distribution of the mean centered about a hypothesized population mean of 500.

The null hypothesis that the population mean for the freshman class equals 500 is *tentatively* assumed to be true. It is tested by determining whether the one observed sample mean qualifies as a common outcome or a rare outcome in the hypothesized sampling distribution of Figure 10.1.

Common Outcomes

An observed sample mean qualifies as a *common* outcome if the difference between its value and that of the hypothesized population mean is small enough to be viewed as a probable outcome under the null hypothesis.

That is, a sample mean qualifies as a common outcome if it doesn't deviate too far from the hypothesized population mean but appears to emerge from the dense concentration of possible sample means in the middle of the sampling distribution. *A common outcome signifies a lack of evidence that, with respect to the null hypothesis, something special is happening in the underlying population*. Because now there is no compelling reason for rejecting the null hypothesis, it is retained.

Key Point:
Does the one observed sample mean qualify as a common or a rare outcome?

Rare Outcomes

An observed sample mean qualifies as a *rare* outcome if the difference between its value and the hypothesized population mean is too large to be reasonably viewed as a probable outcome under the null hypothesis.

That is, a sample mean qualifies as a rare outcome if it deviates too far from the hypothesized mean and appears to emerge from the sparse concentration of possible sample means in either tail of the sampling distribution. *A rare outcome signifies that, with respect to the null hypothesis, something special probably is happening in the underlying population.* Because now there are grounds for suspecting the null hypothesis, it is rejected.

Boundaries for Common and Rare Outcomes

Superimposed on the hypothesized sampling distribution in **Figure 10.2** is one possible set of boundaries for common and rare outcomes, expressed in values of $\overline{X}$. (Techniques for constructing these boundaries are described in Section 10.7.) If the one observed sample mean is located between 478 and 522, it will qualify as a common outcome (readily attributed to variability) under the null hypothesis, and the null

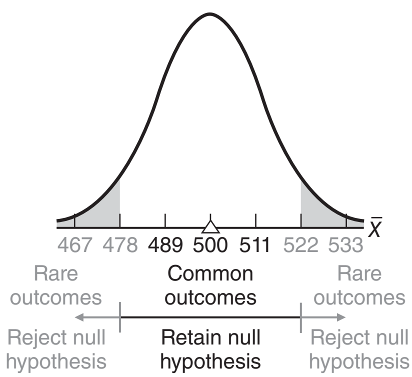

FIGURE 10.2
One possible set of common and rare outcomes (values of $\overline{X}$).

hypothesis will be retained. If, however, the one observed sample mean is greater than 522 or less than 478, it will qualify as a rare outcome (not readily attributed to variability) under the null hypothesis, and the null hypothesis will be rejected. Because the observed sample mean of 533 does exceed 522, the null hypothesis is rejected. On the basis of the present test, it is unlikely that the sample of 100 freshmen, with a mean math score of 533, originates from a population whose mean equals the national average of 500, and, therefore, the investigator can conclude that the mean math score for the local population of freshmen probably differs from (exceeds) the national average.

10.2 *z* TEST FOR A POPULATION MEAN

For the hypothesis test with SAT math scores, it is customary to base the test not on the hypothesized sampling distribution of $\overline{X}$ shown in Figure 10.2, but on its standardized counterpart, the hypothesized sampling distribution of z shown in **Figure 10.3.** Now z represents a variation on the familiar standard score, and it displays all of the properties of standard scores described in Chapter 5. Furthermore, like the sampling distribution of $\overline{X}$, the **sampling distribution of *z*** *represents the distribution of z values that would be obtained if a value of z were calculated for each sample mean for all possible random samples of a given size from some population*.

Sampling Distribution of z

The distribution of z values that would be obtained if a value of z were calculated for each sample mean for all possible random samples of a given size from some population.

The conversion from $\overline{X}$ to z yields a distribution that approximates the standard normal curve in Table A of Appendix C, since, as indicated in Figure 10.3, the original hypothesized population mean (500) emerges as a z score of 0 and the original standard error of the mean (11) emerges as a z score of 1. The shift from $\overline{X}$ to z eliminates the original units of measurement and standardizes the hypothesis test across all situations without, however, affecting the test results.

Reminder: Converting a Raw Score to *z*

To convert a raw score into a standard score (also described in Chapter 5), express the raw score as a distance from its mean (by subtracting the mean from the raw score), and then split this distance into standard deviation units (by dividing with the standard deviation). Expressing this definition as a word formula, we have

$$\textit{Standard score} = \frac{\textit{raw score} - \textit{mean}}{\textit{standard deviation}}$$

in which, of course, the standard score indicates the deviation of the raw score in standard deviation units, above or below the mean.

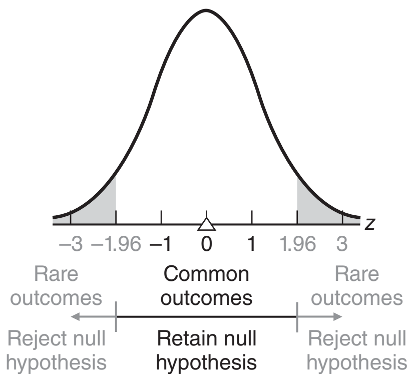

FIGURE 10.3
Common and rare outcomes (values of z).

Converting a Sample Mean to *z*

The z for the present situation emerges as a slight variation of this word formula: Replace the *raw score* with the one observed sample mean $\overline{X}$; replace the *mean* with the mean of the sampling distribution, that is, the hypothesized population mean μ_{hyp}; and replace the *standard deviation* with the standard error of the mean $\sigma_{\overline{x}}$. Now

z RATIO FOR A SINGLE POPULATION MEAN

$$z = \frac{\overline{X} - \mu_{\text{hyp}}}{\sigma_{\overline{x}}} \tag{10.1}$$

where z indicates the deviation of the observed sample mean in standard error units, above or below the hypothesized population mean.

To test the hypothesis for SAT scores, we must determine the value of z from Formula 10.1. Given a sample mean of 533, a hypothesized population mean of 500, and a standard error of 11, we find

$$z = \frac{533 - 500}{11} = \frac{33}{11} = 3$$

The observed z of 3 exceeds the value of 1.96 specified in the hypothesized sampling distribution in Figure 10.3. Thus, the observed z qualifies as a rare outcome under the null hypothesis, and the null hypothesis is rejected. The results of this test with z are the same as those for the original hypothesis test with $\overline{X}$.

Assumptions of *z* Test

When a hypothesis test evaluates how far the observed sample mean deviates, in standard error units, from the hypothesized population mean, as in the present example, *it is referred to as a z test* or, more accurately, as a ***z* test for a population mean.** This z test is accurate only when (1) the population is normally distributed or the sample size is large enough to satisfy the requirements of the central limit theorem and (2) the population standard deviation is known. In the present example, the z test is appropriate because the sample size of 100 is large enough to satisfy the central limit theorem and the population standard deviation is known to be 110.

***z* Test for a Population Mean**
A hypothesis test that evaluates how far the observed sample mean deviates, in standard error units, from the hypothesized population mean.

Progress Check *10.1 Calculate the value of the z test for each of the following situations:

(a) $\overline{X} = 566; \sigma = 30; n = 36; \mu_{hyp} = 560$

(b) $\overline{X} = 24; \sigma = 4; n = 64; \mu_{hyp} = 25$

(c) $\overline{X} = 82; \sigma = 14; n = 49; \mu_{hyp} = 75$

(d) $\overline{X} = 136; \sigma = 15; n = 25; \mu_{hyp} = 146$

Answers on page 432.

10.3 STEP-BY-STEP PROCEDURE

Having been exposed to some of the more important features of hypothesis testing, let's take a detailed look at the test for SAT scores. The test procedure lends itself to a step-by-step description, beginning with a brief statement of the problem that inspired the test and ending with an interpretation of the test results. The following box summarizes the step-by-step procedure for the current hypothesis test. Whenever appropriate, this format will be used in the remainder of the book. Refer to it while reading the remainder of the chapter.

10.4 STATEMENT OF THE RESEARCH PROBLEM

The formulation of a research problem often represents the most crucial and exciting phase of an investigation. Indeed, the mark of a skillful investigator is to focus on an important research problem that can be answered. Do children from broken families score lower on tests of personal adjustment? Do aggressive TV cartoons incite more disruptive behavior in preschool children? Does profit sharing increase the productivity of employees? Because of our emphasis on hypothesis testing, research problems appear in this book as finished products, usually in the first one or two sentences of a new example.

HYPOTHESIS TEST SUMMARY: *z* TEST FOR A POPULATION MEAN (SAT SCORES)

Research Problem

Does the mean SAT math score for all local freshmen differ from the national average of 500?

Statistical Hypotheses

$$H_0: \mu = 500$$
$$H_1: \mu \neq 500$$

Decision Rule

Reject H_0 at the .05 level of significance if $z \geq 1.96$ or if $z \leq -1.96$.

Calculations

Given

$$\overline{X} = 533; \mu_{\text{hyp}} = 500; \sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{110}{\sqrt{100}} = 11$$

$$z = \frac{533 - 500}{11} = 3$$

Decision

Reject H_0 at the .05 level of significance because $z = 3$ exceeds 1.96.

Interpretation

The mean SAT math score for all local freshmen does not equal—it exceeds—the national average of 500.

10.5 NULL HYPOTHESIS (H_0)

Once the problem has been described, it must be translated into a statistical hypothesis regarding some population characteristic. Abbreviated as H_0, the null hypothesis becomes the focal point for the entire test procedure (even though we usually hope to reject it). In the test with SAT scores, the null hypothesis asserts that, with respect to the national average of 500, nothing special is happening to the mean score for the local population of freshmen. An equivalent statement, in symbols, reads:

$$H_0 : \mu = 500$$

where H_0 represents the null hypothesis and μ is the population mean for the local freshman class.

Null Hypothesis (H_0)
A statistical hypothesis that usually asserts that nothing special is happening with respect to some characteristic of the underlying population.

Generally speaking, the **null hypothesis (H_0)** *is a statistical hypothesis that usually asserts that nothing special is happening with respect to some characteristic of the underlying population*. Because the hypothesis testing procedure requires that the hypothesized sampling distribution of the mean be centered about a single number (500), the null hypothesis equals a single number (H_0: μ = 500). Furthermore, the null hypothesis always makes a precise statement about a characteristic of the population, never about a sample. Remember, the purpose of a hypothesis test is to determine whether a particular outcome, such as an observed sample mean, could have reasonably originated from a population with the hypothesized characteristic.

Finding the Single Number for H_0

The single number actually used in H_0 varies from problem to problem. Even for a given problem, this number could originate from any of several sources. For instance, it could be based on available information about some relevant population other than the target population, as in the present example in which 500 reflects the mean SAT math scores for all college-bound students during a recent year. It also could be based on some existing standard or theory—for example, that the mean math score for the current population of local freshmen should equal 540 because that happens to be the mean score achieved by all local freshmen during recent years.

If, as sometimes happens, it's impossible to identify a meaningful null hypothesis, don't try to salvage the situation with arbitrary numbers. Instead, use another entirely different technique, known as *estimation,* which is described in Chapter 12.

10.6 ALTERNATIVE HYPOTHESIS (H_1)

Alternative Hypothesis (H_1)
The opposite of the null hypothesis.

In the present example, the alternative hypothesis asserts that, with respect to the national average of 500, something special is happening to the mean math score for the local population of freshmen (because the mean for the local population doesn't equal the national average of 500). An equivalent statement, in symbols, reads:

$$H_1 : \mu \neq 500$$

where H_1 represents the alternative hypothesis, μ is the population mean for the local freshman class, and ≠ signifies, "is not equal to."

The **alternative hypothesis (H_1)** *asserts the opposite of the null hypothesis.* A decision to retain the null hypothesis implies a lack of support for the alternative hypothesis, and a decision to reject the null hypothesis implies support for the alternative hypothesis.

As will be described in the next chapter, the alternative hypothesis may assume any one of three different forms, depending on the perspective of the investigator. In its present form, H_1 specifies a *range* of possible values about the *single* number (500) that appears in H_0.

Research Hypothesis
Usually identified with the alternative hypothesis, this is the informal hypothesis or hunch that inspires the entire investigation.

Regardless of its form, H_1 usually is identified with the **research hypothesis,** *the informal hypothesis or hunch that, by implying the presence of something special in the underlying population, serves as inspiration for the entire investigation*. "Something special" might be, as in the current example, a deviation from a national average, or it could be, as in later chapters, a deviation from some control condition produced by a new teaching method, a weight-reduction diet, or a self-improvement workshop. In any event, it is this research hypothesis—and certainly not the null hypothesis—that supplies the motive behind an investigation.

Progress Check *10.2 Indicate what's wrong with each of the following statistical hypotheses:

(a) $H_0: \mu = 155$
$H_1: \mu \neq 160$

(b) $H_0: \overline{X} = 241$
$H_1: \overline{X} \neq 241$

Progress Check *10.3 First using words, then symbols, identify the null hypothesis for each of the following situations. (Don't concern yourself about the precise form of the alternative hypothesis at this point.)

(a) A school administrator wishes to determine whether sixth-grade boys in her school district differ, on average, from the national norms of 10.2 pushups for sixth-grade boys.

(b) A consumer group investigates whether, on average, the true weights of packages of ground beef sold by a large supermarket chain differ from the specified 16 ounces.

(c) A marriage counselor wishes to determine whether, during a standard conflict-resolution session, his clients differ, on average, from the 11 verbal interruptions reported for "well-adjusted couples."

Answers on page 432.

10.7 DECISION RULE

Decision Rule
Specifies precisely when H_0 should be rejected (because the observed z qualifies as a rare outcome).

A **decision rule** *specifies precisely when H_0 should be rejected* (because the observed z qualifies as a rare outcome). There are many possible decision rules, as will be seen in Section 11.3. A very common one, already introduced in Figure 10.3, specifies that H_0 should be *rejected* if the observed z equals or is more positive than 1.96 or if the observed z equals or is more negative than –1.96. Conversely, H_0 should be *retained* if the observed z falls between ± 1.96.

Critical *z* Scores

Critical z Score
A z score that separates common from rare outcomes and hence dictates whether H_0 should be retained or rejected.

Figure 10.4 indicates that z scores of ± 1.96 define the boundaries for the middle .95 of the total area (1.00) under the hypothesized sampling distribution for z. Derived from the normal curve table, as you can verify by checking Table A in Appendix C, these two z scores *separate common from rare outcomes and hence dictate whether H_0 should be retained or rejected*. Because of their vital role in the decision about H_0, these scores are referred to as **critical *z* scores.**

Level of Significance (α)

Figure 10.4 also indicates the proportion (.025 + .025 = .05) of the total area that is identified with rare outcomes. Often referred to as the level of significance of the statistical test, this proportion is symbolized by the Greek letter α (alpha) and discussed more thoroughly in Section 11.4. In the present example, the level of significance, α, equals .05.

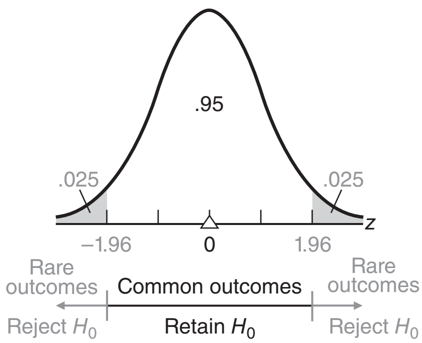

FIGURE 10.4
Proportions of area associated with common and rare outcomes ($\alpha = 05$).

Level of Significance (α)
The degree of rarity required of an observed outcome in order to reject the null hypothesis (H_0).

The **level of significance** (α) *indicates the degree of rarity required of an observed outcome in order to reject the null hypothesis* (H_0). For instance, the .05 level of significance indicates that H_0 should be rejected if the observed z could have occurred just by chance with a probability of only .05 (one chance out of twenty) *or less*.

10.8 CALCULATIONS

We can use information from the sample to calculate a value for z. As has been noted previously, use Formula 10.1 to convert the observed sample mean of 533 into a z of 3.

10.9 DECISION

Either retain or reject H_0, depending on the location of the observed z value relative to the critical z values specified in the decision rule. According to the present rule, H_0 should be rejected at the .05 level of significance because the observed z of 3 exceeds the critical z of 1.96 and, therefore, qualifies as a rare outcome, that is, an unlikely outcome from a population centered about the null hypothesis.

Retain or Reject H_0?

If you are ever confused about whether to retain or reject H_0, recall the logic behind the hypothesis test. You want to reject H_0 only if the observed value of z qualifies as a rare outcome because it deviates too far into the tails of the sampling distribution. Therefore, you want to reject H_0 only if the observed value of z equals or is more positive than the upper critical z (1.96) or if it equals or is more negative than the lower critical z (–1.96). Before deciding, you might find it helpful to sketch the hypothesized sampling distribution, along with its critical z values and shaded rejection regions, and then use some mark, such as an arrow ($\uparrow$), to designate the location of the observed value of z (3) along the z scale. If this mark is located in the shaded rejection region—or farther out than this region, as in Figure 10.4—then H_0 should be rejected.

Progress Check *10.4 For each of the following situations, indicate whether H_0 should be retained or rejected and justify your answer by specifying the precise relationship between observed and critical z scores. Should H_0 be retained or rejected, given a hypothesis test with critical z scores of $\pm$ 1.96 and

(a) $z = 1.74$ **(b)** $z = 0.13$ **(c)** $z = -2.51$

Answers on page 432.

10.10 INTERPRETATION

Finally, interpret the decision in terms of the original research problem. In the present example, it can be concluded that, since the null hypothesis was rejected, the mean SAT math score for the local freshman class probably differs from the national average of 500.

Although not a strict consequence of the present test, a more specific conclusion is possible. Since the sample mean of 533 (or its equivalent z of 3) falls in the *upper* rejection region of the hypothesized sampling distribution, it can be concluded that the population mean SAT math score for all local freshmen probably *exceeds* the national average of 500. By the same token, if the observed sample mean or its equivalent z had fallen in the *lower* rejection region of the hypothesized sampling distribution, it could have been concluded that the population mean for all local freshmen probably is *below* the national average.

If the observed sample mean or its equivalent z had fallen in the retention region of the hypothesized sampling distribution, it would have been concluded (somewhat weakly, as discussed in Section 11.2) that there is no evidence that the population mean for all local freshmen differs from the national average of 500.

Progress Check *10.5 According to the American Psychological Association, members with a doctorate and a full-time teaching appointment earn, on the average, $82,500 per year, with a standard deviation of $6,000. An investigator wishes to determine whether $82,500 is also the mean salary for all female members with a doctorate and a full-time teaching appointment. Salaries are obtained for a random sample of 100 women from this population, and the mean salary equals $80,100.

(a) Someone claims that the observed difference between $80,100 and $82,500 is large enough by itself to support the conclusion that female members earn less than male members. Explain why it is important to conduct a hypothesis test.

(b) The investigator wishes to conduct a hypothesis test for what population?

(c) What is the null hypothesis, H_0?

(d) What is the alternative hypothesis, H_1?

(e) Specify the decision rule, using the .05 level of significance.

(f) Calculate the value of z. (Remember to convert the standard deviation to a standard error.)

(g) What is your decision about H_0?

(h) Using words, interpret this decision in terms of the original problem.

Answers on page 433.

Summary

To test a hypothesis about the population mean, a single observed sample mean is viewed within the context of a hypothesized sampling distribution, itself centered about the null-hypothesized population mean. If the sample mean appears to emerge from the dense concentration of possible sample means in the middle of the sampling distribution, it qualifies as a common outcome, and the null hypothesis is retained. On the other hand, if the sample mean appears to emerge from the sparse concentration of possible *sample* means at the extremities of the *sampling* distribution, it qualifies as a rare outcome, and the null hypothesis is rejected.

Hypothesis tests are based not on the sampling distribution of $\overline{X}$ expressed in original units of measurement, but on its standardized counterpart, the sampling distribution of z. Referred to as the z test for a single population mean, this test is appropriate only when (1) the population is normally distributed or the sample size is large enough to satisfy the central limit theorem, and (2) the population standard deviation is known.

When testing a hypothesis, adopt the following step-by-step procedure:

- **State the research problem.** Using words, state the problem to be resolved by the investigation.
- **Identify the statistical hypotheses.** The statistical hypotheses consist of a null hypothesis (H_0) and an alternative (or research) hypothesis (H_1). The null hypothesis supplies the value about which the hypothesized sampling distribution is centered. Depending on the outcome of the hypothesis test, H_0 will either be retained or rejected. Insofar as H_0 implies that nothing special is happening in the underlying population, the investigator usually hopes to reject it in favor of H_1, the research hypothesis. In the present chapter, the statistical hypotheses take the form

$$H_0 : \mu = \text{some number}$$
$$H_1 : \mu \neq \text{some number}$$

(Two other possible forms for statistical hypotheses will be described in Chapter 11.)

- **Specify a decision rule.** This rule indicates precisely when H_0 should be rejected. The exact form of the decision rule depends on a number of factors, to be discussed in Chapter 11. In any event, H_0 is rejected whenever the observed z deviates from 0 as far as, or farther than, the critical z does.

  The level of significance indicates how rare an observed z must be (assuming that H_0 is true) before H_0 can be rejected.

- **Calculate the value of the observed z.** Express the one observed sample mean as an observed z, using Formula 10.1.
- **Make a decision.** Either retain or reject H_0 at the specified level of significance, justifying this decision by noting the relationship between observed and critical z scores.
- **Interpret the decision**. Using words, interpret the decision in terms of the original research problem. Rejection of the null hypothesis supports the research hypothesis, while retention of the null hypothesis fails to support the research hypothesis.

Important Terms

Sampling distribution of *z*
Null hypothesis (H_0)
Research hypothesis
Critical *z* score
***z* Test for a population mean**
Alternative hypothesis (H_1)
Decision rule
Level of significance (α)

Key Equations

***z* RATIO**

$$z = \frac{\overline{X} - \mu_{\text{hyp}}}{\sigma_{\overline{x}}}$$

$$\text{where } \sigma_{\overline{x}} = \frac{\sigma}{\sqrt{n}}$$

REVIEW QUESTIONS

10.6 Calculate the value of the z test for each of the following situations.

(a) $\overline{X} = 12$; $\sigma = 9$; $n = 25$, $\mu_{hyp} = 15$

(b) $\overline{X} = 3600$; $\sigma = 4000$; $n = 100$; $\mu_{hyp} = 3500$

(c) $\overline{X} = 0.25$; $\sigma = 010$; $n = 36$; $\mu_{hyp} = 0.22$

10.7 Given critical z scores of ± 1.96, should H_0 be accepted or rejected for each of the z scores calculated in Exercise 10.6?

***10.8** For the population at large, the Wechsler Adult Intelligence Scale is designed to yield a normal distribution of test scores with a mean of 100 and a standard deviation of 15. School district officials wonder whether, on the average, an IQ score different from 100 describes the intellectual aptitudes of all students in their district. Wechsler IQ scores are obtained for a random sample of 25 of their students, and the mean IQ is found to equal 105. Using the step-by-step procedure described in this chapter, test the null hypothesis at the .05 level of significance.

Answers on page 433.

10.9 The normal range for a widely accepted measure of body size, the body mass index (BMI), ranges from 18.5 to 25. Using the midrange BMI score of 21.75 as the null hypothesized value for the population mean, test this hypothesis at the .01 level of significance given a random sample of 30 weight-watcher participants who show a mean BMI = 22.2 and a standard deviation of 3.1.

10.10 Let's assume that, over the years, a paper and pencil test of anxiety yields a mean score of 35 for all incoming college freshmen. We wish to determine whether the scores of a random sample of 20 new freshmen, with a mean of 30 and a standard deviation of 10, can be viewed as coming from this population. Test at the .05 level of significance.

10.11 According to the California Educational Code (http://www.cde.ca.gov/ls/fa/sf/peguidemidhi.asp), students in grades 7 through 12 should receive 400 minutes of physical education every 10 school days. A random sample of 48 students has a mean of 385 minutes and a standard deviation of 53 minutes. Test the hypothesis at the .05 level of significance that the sampled population satisfies the requirement.

10.12 According to a 2009 survey based on the United States census (http://www.census.gov/prod/2011pubs/acs-15.pdf), the daily one-way commute time of U.S. workers averages 25 minutes with, we'll assume, a standard deviation of 13 minutes. An investigator wishes to determine whether the national average describes the mean commute time for all workers in the Chicago area. Commute times are obtained for a random sample of 169 workers from this area, and the mean time is found to be 22.5 minutes. Test the null hypothesis at the .05 level of significance.

10.13 Supply the missing word(s) in the following statements:

If the one observed sample mean can be viewed as a __(a)__ outcome under the hypothesis, H_0 will be __(b)__. Otherwise, if the one observed sample mean can be viewed as a __(c)__ outcome under the hypothesis, H_0 will be __(d)__.

The pair of *z* scores that separates common and rare outcomes is referred to as __(e)__ *z* scores. Within the hypothesized sampling distribution, the proportion of area allocated to rare outcomes is referred to as the __(f)__ and is symbolized by the Greek letter __(g)__ .

When based on the sampling distribution of *z*, the hypothesis test is referred to as a __(h)__ test. This test is appropriate if the sample size is sufficiently large to satisfy the __(i)__ and if the __(j)__ is known.