CHAPTER

7

© Steve Cole/iStockphoto

Sampling Distributions

You read about opinion polls in newspapers, magazines, and on the Web every day. These polls are based on sample surveys. Have you heard of sampling and nonsampling errors? It is good to be aware of such errors while reading these opinion poll results. Sound sampling methods are essential for opinion poll results to be valid and to lower the effects of such errors.

Chapters 5 and 6 discussed probability distributions of discrete and continuous random variables. This chapter extends the concept of probability distribution to that of a sample statistic. As we discussed in Chapter 3, a sample statistic is a numerical summary measure calculated for sample data. The mean, median, mode, and standard deviation calculated for sample data are called *sample statistics*. On the other hand, the same numerical summary measures calculated for population data are called *population parameters*. A population parameter is always a constant (at a given point in time), whereas a sample statistic is always a random variable. Because every random variable must possess a probability distribution, each sample statistic possesses a probability distribution. The probability distribution of a sample statistic is more commonly called its *sampling distribution*. This chapter discusses the sampling distributions of the sample mean and the sample proportion. The concepts covered in this chapter are the foundation of the inferential statistics discussed in succeeding chapters.

7.1 Sampling Distribution, Sampling Error, and Nonsampling Errors

This section introduces the concepts of sampling distribution, sampling error, and nonsampling errors. Before we discuss these concepts, we will briefly describe the concept of a population distribution.

The **population distribution** is the probability distribution derived from the information on all elements of a population.

Definition

Population Distribution The *population distribution* is the probability distribution of the population data.

Suppose there are only five students in an advanced statistics class and the midterm scores of these five students are

70 78 80 80 95

Let x denote the score of a student. Using single-valued classes (because there are only five data values, there is no need to group them), we can write the frequency distribution of scores as in Table 7.1 along with the relative frequencies of classes, which are obtained by dividing the frequencies of classes by the population size. Table 7.2, which lists the probabilities of various x values, presents the probability distribution of the population. Note that these probabilities are the same as the relative frequencies.

Table 7.1 Population Frequency and Relative Frequency Distributions

x	f	Relative Frequency
70	1	1/5 = .20
78	1	1/5 = .20
80	2	2/5 = .40
95	1	1/5 = .20
	$N = 5$	Sum = 1.00

Table 7.2 Population Probability Distribution

x	$P(x)$
70	.20
78	.20
80	.40
95	.20
	$\Sigma P(x) = 1.00$

The values of the mean and standard deviation calculated for the probability distribution of Table 7.2 give the values of the population parameters μ and σ. These values are $\mu = 80.60$ and $\sigma = 8.09$. The values of μ and σ for the probability distribution of Table 7.2 can be calculated using the formulas given in Section 5.3 of Chapter 5 (see Exercise 7.6).

7.1.1 Sampling Distribution

As mentioned at the beginning of this chapter, the value of a population parameter is always constant. For example, for any population data set, there is only one value of the population mean, μ. However, we cannot say the same about the sample mean, $\bar{x}$. We would expect different samples of the same size drawn from the same population to yield different values of the sample mean, $\bar{x}$. The value of the sample mean for any one sample will depend on the elements included in that sample. Consequently, *the sample mean*, $\bar{x}$, *is a random variable*. Therefore, like other random variables, the sample mean possesses a probability distribution, which is more commonly called the **sampling distribution of $\bar{x}$**. Other sample statistics, such as the median, mode, and standard deviation, also possess sampling distributions.

Definition

Sampling Distribution of $\bar{x}$ The probability distribution of $\bar{x}$ is called its sampling distribution. It lists the various values that $\bar{x}$ can assume and the probability of each value of $\bar{x}$.

In general, the probability distribution of a sample statistic is called its *sampling distribution*.

Reconsider the population of midterm scores of five students given in Table 7.1. Consider all possible samples of three scores each that can be selected, without replacement, from that population. The total number of possible samples, given by the combinations formula discussed in Chapter 4, is 10; that is,

$$\text{Total number of samples} = {}_5C_3 = \frac{5!}{3!(5-3)!} = \frac{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{3 \cdot 2 \cdot 1 \cdot 2 \cdot 1} = 10$$

Suppose we assign the letters A, B, C, D, and E to the scores of the five students, so that

$$\text{A} = 70, \quad \text{B} = 78, \quad \text{C} = 80, \quad \text{D} = 80, \quad \text{E} = 95$$

Then, the 10 possible samples of three scores each are

ABC, ABD, ABE, ACD, ACE, ADE, BCD, BCE, BDE, CDE

These 10 samples and their respective means are listed in Table 7.3. Note that the first two samples have the same three scores. The reason for this is that two of the students (C and D) have the same score, and, hence, the samples ABC and ABD contain the same values. The mean of each sample is obtained by dividing the sum of the three scores included in that sample by 3. For instance, the mean of the first sample is $(70 + 78 + 80)/3 = 76$. Note that the values of the means of samples in Table 7.3 are rounded to two decimal places.

By using the values of $\bar{x}$ given in Table 7.3, we record the frequency distribution of $\bar{x}$ in Table 7.4. By dividing the frequencies of the various values of $\bar{x}$ by the sum of all frequencies, we obtain the relative frequencies of classes, which are listed in the third column of Table 7.4. These relative frequencies are used as probabilities and listed in Table 7.5. This table gives the sampling distribution of $\bar{x}$.

If we select just one sample of three scores from the population of five scores, we may draw any of the 10 possible samples. Hence, the sample mean, $\bar{x}$, can assume any of the values listed in Table 7.5 with the corresponding probability. For instance, the probability that the mean of a randomly selected sample of three scores is 81.67 is .20. This probability can be written as

$$P(\bar{x} = 81.67) = .20$$

Table 7.3 All Possible Samples and Their Means When the Sample Size Is 3

Sample	Scores in the Sample	$\bar{x}$
ABC	70, 78, 80	76.00
ABD	70, 78, 80	76.00
ABE	70, 78, 95	81.00
ACD	70, 80, 80	76.67
ACE	70, 80, 95	81.67
ADE	70, 80, 95	81.67
BCD	78, 80, 80	79.33
BCE	78, 80, 95	84.33
BDE	78, 80, 95	84.33
CDE	80, 80, 95	85.00

Table 7.4 Frequency and Relative Frequency Distributions of $\bar{x}$ When the Sample Size Is 3

$\bar{x}$	f	Relative Frequency
76.00	2	2/10 = .20
76.67	1	1/10 = .10
79.33	1	1/10 = .10
81.00	1	1/10 = .10
81.67	2	2/10 = .20
84.33	2	2/10 = .20
85.00	1	1/10 = .10
	$\Sigma f = 10$	Sum = 1.00

Table 7.5 Sampling Distribution of $\bar{x}$ When the Sample Size Is 3

$\bar{x}$	$P(\bar{x})$
76.00	.20
76.67	.10
79.33	.10
81.00	.10
81.67	.20
84.33	.20
85.00	.10
	$\Sigma P(\bar{x}) = 1.00$

7.1.2 Sampling and Nonsampling Errors

Usually, different samples selected from the same population will give different results because they contain different elements. This is obvious from Table 7.3, which shows that the mean of a sample of three scores depends on which three of the five scores are included in the sample. The result obtained from any one sample will generally be different from the result obtained from the corresponding population. The difference between the value of a sample statistic obtained from a sample and the value of the corresponding population parameter obtained from the population is called the **sampling error**. Note that this difference represents the sampling error only if the sample is random and no nonsampling error has been made. Otherwise, only a part of this difference will be due to the sampling error.

Definition

Sampling Error *Sampling error* is the difference between the value of a sample statistic and the value of the corresponding population parameter. In the case of the mean,

$$\text{Sampling error} = \bar{x} - \mu$$

assuming that the sample is random and no nonsampling error has been made.

It is important to remember that *a sampling error occurs because of chance*. The errors that occur for other reasons, such as errors made during collection, recording, and tabulation of data, are called **nonsampling errors**. These errors occur because of human mistakes, and not chance. Note that there is only one kind of sampling error—the error that occurs due to chance. However, there is not just one nonsampling error, but there are many nonsampling errors that may occur for different reasons.

Definition

Nonsampling Errors The errors that occur in the collection, recording, and tabulation of data are called *nonsampling errors*.

The following paragraph, reproduced from the *Current Population Reports* of the U.S. Bureau of the Census, explains how nonsampling errors can occur.

> Nonsampling errors can be attributed to many sources, e.g., inability to obtain information about all cases in the sample, definitional difficulties, differences in the interpretation of questions, inability or unwillingness on the part of the respondents to provide correct information, inability to recall information, errors made in collection such as in recording or coding the data, errors made in processing the data, errors made in estimating values for missing data, biases resulting from the differing recall periods caused by the interviewing pattern used, and failure of all units in the universe to have some probability of being selected for the sample (undercoverage).

The following are the main reasons for the occurrence of nonsampling errors.

1. If a sample is nonrandom (and, hence, most likely nonrepresentative), the sample results may be too different from the census results. Even a randomly selected sample can become nonrandom if some of the members included in the sample cannot be contacted. A very good example of this comes from an article published in a magazine in 1988. As reported in a July 11, 1988, article in *U.S. News & World Report* ("The Numbers Racket: How Polls and Statistics Lie"), during the 1984 presidential election a test poll was conducted in which the only subjects interviewed were those who could be reached on the first try. The results of this poll indicated that Ronald Reagan had a 3 percentage point lead over Walter Mondale. However, when interviewers made an effort to contact everyone on their lists (calling some households up to 30 times before reaching someone), this lead increased to 13%. It turned out that this 13% lead was much closer to the actual election results. Apparently, people who planned to vote Republican spent less time at home.
2. The questions may be phrased in such a way that they are not fully understood by the members of the sample or population. As a result, the answers obtained are not accurate.
3. The respondents may intentionally give false information in response to some sensitive questions. For example, people may not tell the truth about their drinking habits, incomes, or opinions about minorities. Sometimes the respondents may give wrong answers because of ignorance. For example, a person may not remember the exact amount he or she spent on clothes last year. If asked in a survey, he or she may give an inaccurate answer.
4. The poll taker may make a mistake and enter a wrong number in the records or make an error while entering the data on a computer.

Note that nonsampling errors can occur both in a sample survey and in a census, whereas sampling error occurs only when a sample survey is conducted. Nonsampling errors can be minimized by preparing the survey questionnaire carefully and handling the data cautiously. However, it is impossible to avoid sampling error.

Example 7–1 illustrates the sampling and nonsampling errors using the mean.

■ EXAMPLE 7–1

Illustrating sampling and nonsampling errors.

Reconsider the population of five scores given in Table 7.1. Suppose one sample of three scores is selected from this population, and this sample includes the scores 70, 80, and 95. Find the sampling error.

Solution The scores of the five students are 70, 78, 80, 80, and 95. The population mean is

$$\mu = \frac{70 + 78 + 80 + 80 + 95}{5} = 80.60$$

Now a random sample of three scores from this population is taken and this sample includes the scores 70, 80, and 95. The mean for this sample is

$$\bar{x} = \frac{70 + 80 + 95}{3} = 81.67$$

Consequently,

$$\text{Sampling error} = \bar{x} - \mu = 81.67 - 80.60 = \mathbf{1.07}$$

That is, the mean score estimated from the sample is 1.07 higher than the mean score of the population. Note that this difference occurred due to chance—that is, because we used a sample instead of the population. ■

Now suppose, when we select the sample of three scores, we mistakenly record the second score as 82 instead of 80. As a result, we calculate the sample mean as

$$\bar{x} = \frac{70 + 82 + 95}{3} = 82.33$$

Consequently, the difference between this sample mean and the population mean is

$$\bar{x} - \mu = 82.33 - 80.60 = 1.73$$

However, this difference between the sample mean and the population mean does not represent the sampling error. As we calculated earlier, only 1.07 of this difference is due to the sampling error. The remaining portion, which is equal to $1.73 - 1.07 = .66$, represents the nonsampling error because it occurred due to the error we made in recording the second score in the sample. Thus, in this case,

$$\text{Sampling error} = \mathbf{1.07}$$

$$\text{Nonsampling error} = \mathbf{.66}$$

Figure 7.1 shows the sampling and nonsampling errors for these calculations.

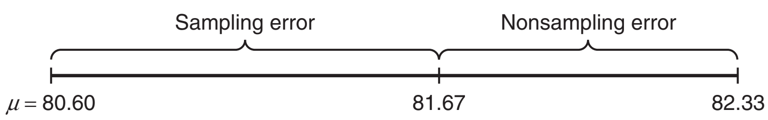


Figure 7.1 Sampling and nonsampling errors.

Thus, the sampling error is the difference between the correct value of $\bar{x}$ and the value of μ, where the correct value of $\bar{x}$ is the value of $\bar{x}$ that does not contain any nonsampling errors. In contrast, the nonsampling error(s) is (are) obtained by subtracting the correct value of $\bar{x}$ from the incorrect value of $\bar{x}$, where the incorrect value of $\bar{x}$ is the value that contains the nonsampling error(s). For our example,

$$\text{Sampling error} = \bar{x} - \mu = 81.67 - 80.60 = 1.07$$

$$\text{Nonsampling error} = \text{Incorrect } \bar{x} - \text{Correct } \bar{x} = 82.33 - 81.67 = .66$$

Note that in the real world we do not know the mean of a population. Hence, we select a sample to use the sample mean as an estimate of the population mean. Consequently, we never know the size of the sampling error.

EXERCISES

CONCEPTS AND PROCEDURES

7.1 Briefly explain the meaning of a population distribution and a sampling distribution. Give an example of each.

7.2 Explain briefly the meaning of sampling error. Give an example. Does such an error occur only in a sample survey, or can it occur in both a sample survey and a census?

7.3 Explain briefly the meaning of nonsampling errors. Give an example. Do such errors occur only in a sample survey, or can they occur in both a sample survey and a census?

7.4 Consider the following population of six numbers.

16	13	7	20	9	12

a. Find the population mean.
b. Liza selected one sample of four numbers from this population. The sample included the numbers 13, 7, 9, and 12. Calculate the sample mean and sampling error for this sample.
c. Refer to part b. When Liza calculated the sample mean, she mistakenly used the numbers 13, 7, 6, and 12 to calculate the sample mean. Find the sampling and nonsampling errors in this case.
d. List all samples of four numbers (without replacement) that can be selected from this population. Calculate the sample mean and sampling error for each of these samples.

7.5 Consider the following population of 10 numbers.

20	18	13	12	9	5	11	7	8	30

a. Find the population mean.
b. Rich selected one sample of nine numbers from this population. The sample included the numbers 20, 18, 13, 9, 5, 11, 7, 8, and 30. Calculate the sample mean and sampling error for this sample.
c. Refer to part b. When Rich calculated the sample mean, he mistakenly used the numbers 20, 18, 13, 9, 5, 11, 8, 8, and 30 to calculate the sample mean. Find the sampling and nonsampling errors in this case.
d. List all samples of nine numbers (without replacement) that can be selected from this population. Calculate the sample mean and sampling error for each of these samples.

■ APPLICATIONS

7.6 Using the formulas of Section 5.3 of Chapter 5 for the mean and standard deviation of a discrete random variable, verify that the mean and standard deviation for the population probability distribution of Table 7.2 are 80.60 and 8.09, respectively.

7.7 The following data give the ages (in years) of all five members of a family.

55	53	28	25	21

a. Let x denote the age of a member of this family. Write the population distribution of x.
b. List all the possible samples of size five (without replacement) that can be selected from this population. Calculate the mean for each of these samples. Write the sampling distribution of $\bar{x}$.
c. Calculate the mean for the population data. Select one random sample of size four and calculate the sample mean $\bar{x}$. Compute the sampling error.

7.8 The following data give the years of teaching experience for all six faculty members of a department at a university.

7	8	14	7	20	12

a. Let x denote the years of teaching experience for a faculty member of this department. Write the population distribution of x.
b. List all the possible samples of size five (without replacement) that can be selected from this population. Calculate the mean for each of these samples. Write the sampling distribution of $\bar{x}$.
c. Calculate the mean for the population data. Select one random sample of size five and calculate the sample mean $\bar{x}$. Compute the sampling error.

7.2 Mean and Standard Deviation of $\bar{x}$

The mean and standard deviation calculated for the sampling distribution of $\bar{x}$ are called the **mean** and **standard deviation of $\bar{x}$**. Actually, the mean and standard deviation of $\bar{x}$ are, respectively, the mean and standard deviation of the means of all samples of the same size selected from a population. The standard deviation of $\bar{x}$ is also called the **standard error of $\bar{x}$**.

Definition

Mean and Standard Deviation of $\bar{x}$ The mean and standard deviation of the sampling distribution of $\bar{x}$ are called the *mean and standard deviation of* $\bar{x}$ and are denoted by $\mu_{\bar{x}}$ and $\sigma_{\bar{x}}$, respectively.

If we calculate the mean and standard deviation of the 10 values of $\bar{x}$ listed in Table 7.3, we obtain the mean, $\mu_{\bar{x}}$, and the standard deviation, $\sigma_{\bar{x}}$, of $\bar{x}$. Alternatively, we can calculate the mean and standard deviation of the sampling distribution of $\bar{x}$ listed in Table 7.5. These will also be the values of $\mu_{\bar{x}}$ and $\sigma_{\bar{x}}$. From these calculations, we will obtain $\mu_{\bar{x}} = 80.60$ and $\sigma_{\bar{x}} = 3.30$ (see Exercise 7.25 at the end of this section).

The mean of the sampling distribution of $\bar{x}$ is always equal to the mean of the population.

Mean of the Sampling Distribution of $\bar{x}$ The *mean of the sampling distribution of* $\bar{x}$ is always equal to the mean of the population. Thus,

$$\mu_{\bar{x}} = \mu$$

Thus, if we select all possible samples (of the same size) from a population and calculate their means, the mean ($\mu_{\bar{x}}$) of all these sample means will be the same as the mean (μ) of the population. If we calculate the mean for the population probability distribution of Table 7.2 and the mean for the sampling distribution of Table 7.5 by using the formula learned in Section 5.3 of Chapter 5, we get the same value of 80.60 for μ and $\mu_{\bar{x}}$ (see Exercise 7.25).

The sample mean, $\bar{x}$, is called an **estimator** of the population mean, μ. When the expected value (or mean) of a sample statistic is equal to the value of the corresponding population parameter, that sample statistic is said to be an **unbiased estimator**. For the sample mean $\bar{x}$, $\mu_{\bar{x}} = \mu$. Hence, $\bar{x}$ is an unbiased estimator of μ. This is a very important property that an estimator should possess.

However, the standard deviation, $\sigma_{\bar{x}}$, of $\bar{x}$ is not equal to the standard deviation, σ, of the population distribution (unless $n = 1$). The standard deviation of $\bar{x}$ is equal to the standard deviation of the population divided by the square root of the sample size; that is,

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

This formula for the standard deviation of $\bar{x}$ holds true only when the sampling is done either with replacement from a finite population or with or without replacement from an infinite population. These two conditions can be replaced by the condition that the above formula holds true if the sample size is small in comparison to the population size. The sample size is considered to be small compared to the population size if the sample size is equal to or less than 5% of the population size; that is, if

$$\frac{n}{N} \leq .05$$

If this condition is not satisfied, we use the following formula to calculate $\sigma_{\bar{x}}$:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

where the factor $\sqrt{\frac{N-n}{N-1}}$ is called the finite population correction factor.

In most practical applications, the sample size is small compared to the population size. Consequently, in most cases, the formula used to calculate $\sigma_{\bar{x}}$ is $\sigma_{\bar{x}} = \sigma/\sqrt{n}$.

Standard Deviation of the Sampling Distribution of $\bar{x}$ The *standard deviation of the sampling distribution of* $\bar{x}$ is

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

where σ is the standard deviation of the population and n is the sample size. This formula is used when $n/N \leq .05$, where N is the population size.

Following are two important observations regarding the sampling distribution of $\bar{x}$.

1. *The spread of the sampling distribution of $\bar{x}$ is smaller than the spread of the corresponding population distribution*. In other words, $\sigma_{\bar{x}} < \sigma$. This is obvious from the formula for $\sigma_{\bar{x}}$. When n is greater than 1, which is usually true, the denominator in $\sigma/\sqrt{n}$ is greater than 1. Hence, $\sigma_{\bar{x}}$ is smaller than σ.
2. *The standard deviation of the sampling distribution of $\bar{x}$ decreases as the sample size increases*. This feature of the sampling distribution of $\bar{x}$ is also obvious from the formula

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

If the standard deviation of a sample statistic decreases as the sample size is increased, that statistic is said to be a **consistent estimator**. This is another important property that an estimator should possess. It is obvious from the above formula for $\sigma_{\bar{x}}$ that as n increases, the value of $\sqrt{n}$ also increases and, consequently, the value of $\sigma/\sqrt{n}$ decreases. Thus, the sample mean $\bar{x}$ is a consistent estimator of the population mean μ. Example 7–2 illustrates this feature.

■ EXAMPLE 7–2

Finding the mean and standard deviation of $\bar{x}$.

The mean wage per hour for all 5000 employees who work at a large company is \$27.50, and the standard deviation is \$3.70. Let $\bar{x}$ be the mean wage per hour for a random sample of certain employees selected from this company. Find the mean and standard deviation of $\bar{x}$ for a sample size of

(a) 30 **(b)** 75 **(c)** 200

Image Source/GettyImages, Inc.

Solution From the given information, for the population of all employees,

$$N = 5000, \quad \mu = \$27.50, \quad \text{and} \quad \sigma = \$3.70$$

(a) The mean, $\mu_{\bar{x}}$, of the sampling distribution of $\bar{x}$ is

$$\mu_{\bar{x}} = \mu = \mathbf{\$27.50}$$

In this case, $n = 30$, $N = 5000$, and $n/N = 30/5000 = .006$. Because n/N is less than .05, the standard deviation of $\bar{x}$ is obtained by using the formula $\sigma/\sqrt{n}$. Hence,

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{3.70}{\sqrt{30}} = \mathbf{\$.676}$$

Thus, we can state that if we take all possible samples of size 30 from the population of all employees of this company and prepare the sampling distribution of $\bar{x}$, the mean and standard deviation of this sampling distribution of $\bar{x}$ will be \$27.50 and \$.676, respectively.

(b) In this case, $n = 75$ and $n/N = 75/5000 = .015$, which is less than .05. The mean and standard deviation of $\bar{x}$ are

$$\mu_{\bar{x}} = \mu = \mathbf{\$27.50} \quad \text{and} \quad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{3.70}{\sqrt{75}} = \mathbf{\$.427}$$

(c) In this case, $n = 200$ and $n/N = 200/5000 = .04$, which is less than .05. Therefore, the mean and standard deviation of $\bar{x}$ are

$$\mu_{\bar{x}} = \mu = \mathbf{\$27.50} \quad \text{and} \quad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{3.70}{\sqrt{200}} = \mathbf{\$.262}$$

From the preceding calculations we observe that the mean of the sampling distribution of $\bar{x}$ is always equal to the mean of the population whatever the size of the sample. However, the value of the standard deviation of $\bar{x}$ decreases from \$.676 to \$.427 and then to \$.262 as the sample size increases from 30 to 75 and then to 200. ■

EXERCISES

CONCEPTS AND PROCEDURES

7.9 Let $\bar{x}$ be the mean of a sample selected from a population.

a. What is the mean of the sampling distribution of $\bar{x}$ equal to?

b. What is the standard deviation of the sampling distribution of $\bar{x}$ equal to? Assume $n/N \le .05$.

7.10 What is an estimator? When is an estimator unbiased? Is the sample mean, $\bar{x}$, an unbiased estimator of μ? Explain.

7.11 When is an estimator said to be consistent? Is the sample mean, $\bar{x}$, a consistent estimator of μ? Explain.

7.12 How does the value of $\sigma_{\bar{x}}$ change as the sample size increases? Explain.

7.13 Consider a large population with $\mu = 70$ and $\sigma = 10$. Assuming $n/N \le .05$, find the mean and standard deviation of the sample mean, $\bar{x}$, for a sample size of

a. 18 **b.** 80

7.14 Consider a large population with $\mu = 90$ and $\sigma = 18$. Assuming $n/N \le .05$, find the mean and standard deviation of the sample mean, $\bar{x}$, for a sample size of

a. 15 **b.** 40

7.15 A population of $N = 5000$ has $\sigma = 25$. In each of the following cases, which formula will you use to calculate $\sigma_{\bar{x}}$ and why? Using the appropriate formula, calculate $\sigma_{\bar{x}}$ for each of these cases.

a. $n = 300$ **b.** $n = 100$

7.16 A population of $N = 100{,}000$ has $\sigma = 50$. In each of the following cases, which formula will you use to calculate $\sigma_{\bar{x}}$ and why? Using the appropriate formula, calculate $\sigma_{\bar{x}}$ for each of these cases.

a. $n = 2500$ **b.** $n = 7000$

***7.17** For a population, $\mu = 125$ and $\sigma = 36$.

a. For a sample selected from this population, $\mu_{\bar{x}} = 125$ and $\sigma_{\bar{x}} = 3.6$. Find the sample size. Assume $n/N \le .05$.

b. For a sample selected from this population, $\mu_{\bar{x}} = 125$ and $\sigma_{\bar{x}} = 2.25$. Find the sample size. Assume $n/N \le .05$.

***7.18** For a population, $\mu = 46$ and $\sigma = 10$.

a. For a sample selected from this population, $\mu_{\bar{x}} = 46$ and $\sigma_{\bar{x}} = 2.0$. Find the sample size. Assume $n/N \le .05$.

b. For a sample selected from this population, $\mu_{\bar{x}} = 46$ and $\sigma_{\bar{x}} = 1.6$. Find the sample size. Assume $n/N \le .05$.

APPLICATIONS

7.19 According to the Project on Student Debt, the average student loan for college graduates of the class of 2010 was $25,000 (*USA TODAY*, April 24, 2012). Suppose that the student loans for all college graduates of the class of 2010 have a mean of $25,000 and a standard deviation of $6280. Let $\bar{x}$ be the average student loan of a random sample of 400 college graduates from the class of 2010. Find the mean and standard deviation of the sampling distribution of $\bar{x}$.

7.20 The living spaces of all homes in a city have a mean of 2300 square feet and a standard deviation of 500 square feet. Let $\bar{x}$ be the mean living space for a random sample of 25 homes selected from this city. Find the mean and standard deviation of the sampling distribution of $\bar{x}$.

7.21 According to a report in *The New York Times*, bank tellers in the United States earn an average of $25,510 a year (Jessica Silver-Greenberg, *The New York Times*, April 22, 2012). Suppose that the current distribution of salaries of all bank tellers in the United States has a mean of $25,510 and a standard deviation of $4550. Let $\bar{x}$ be the average salary of a random sample of 200 such tellers. Find the mean and standard deviation of the sampling distribution of $\bar{x}$.

7.22 According to the American Automobile Association's 2012 annual report *Your Driving Costs*, the cost of owning and operating a four-wheel drive SUV is $11,350 per year (*USA TODAY*, April 27, 2012). Note that this cost includes expenses for gasoline, maintenance, insurance, and financing for a vehicle that is driven 15,000 miles a year. Suppose that the distribution of such costs of owning and operating all four-wheel drive SUVs has a mean of $11,350 with a standard deviation of $2390. Let $\bar{x}$ be the average of such costs of owning and operating a four-wheel drive SUV based on a random sample of 400 four-wheel drive SUVs. Find the mean and standard deviation of the sampling distribution of $\bar{x}$.

***7.23** Suppose the standard deviation of recruiting costs per player for all female basketball players recruited by all public universities in the Midwest is $2000. Let $\bar{x}$ be the mean recruiting cost for a sample of a certain number of such players. What sample size will give the standard deviation of $\bar{x}$ equal to $125? Assume $n/N \leq .05$.

***7.24** The standard deviation of the 2011 gross sales of all corporations is known to be $139.50 million. Let $\bar{x}$ be the mean of the 2011 gross sales of a sample of corporations. What sample size will produce the standard deviation of $\bar{x}$ equal to $15 million? Assume $n/N \leq .05$.

***7.25** Consider the sampling distribution of $\bar{x}$ given in Table 7.5.

a. Calculate the value of $\mu_{\bar{x}}$ using the formula $\mu_{\bar{x}} = \Sigma \bar{x} P(\bar{x})$. Is the value of μ calculated in Exercise 7.6 the same as the value of $\mu_{\bar{x}}$ calculated here?

b. Calculate the value of $\sigma_{\bar{x}}$ by using the formula

$$\sigma_{\bar{x}} = \sqrt{\Sigma \bar{x}^2 P(\bar{x}) - (\mu_{\bar{x}})^2}$$

c. From Exercise 7.6, $\sigma = 8.09$. Also, our sample size is 3, so that $n = 3$. Therefore, $\sigma/\sqrt{n} = 8.09/\sqrt{3} = 4.67$. From part b, you should get $\sigma_{\bar{x}} = 3.30$. Why does $\sigma/\sqrt{n}$ not equal $\sigma_{\bar{x}}$ in this case?

d. In our example (given in the beginning of Section 7.1) on scores, $N = 5$ and $n = 3$. Hence, $n/N = 3/5 = .60$. Because n/N is greater than .05, the appropriate formula to find $\sigma_{\bar{x}}$ is

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$$

Show that the value of $\sigma_{\bar{x}}$ calculated by using this formula gives the same value as the one calculated in part b above.

7.3 Shape of the Sampling Distribution of $\bar{x}$

The shape of the sampling distribution of $\bar{x}$ relates to the following two cases:

1. The population from which samples are drawn has a normal distribution.
2. The population from which samples are drawn does not have a normal distribution.

7.3.1 Sampling from a Normally Distributed Population

When the population from which samples are drawn is normally distributed with its mean equal to μ and standard deviation equal to σ, then:

1. The mean of $\bar{x}$, $\mu_{\bar{x}}$, is equal to the mean of the population, μ.
2. The standard deviation of $\bar{x}$, $\sigma_{\bar{x}}$, is equal to $\sigma/\sqrt{n}$, assuming $n/N \leq .05$.
3. The shape of the sampling distribution of $\bar{x}$ is normal, whatever the value of n.

Sampling Distribution of $\bar{x}$ When the Population Has a Normal Distribution If the population from which the samples are drawn is normally distributed with mean μ and standard deviation σ, then the sampling distribution of the sample mean, $\bar{x}$, will also be normally distributed with the following mean and standard deviation, irrespective of the sample size:

$$\mu_{\bar{x}} = \mu \quad \text{and} \quad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

Remember ▶ For $\sigma_{\bar{x}} = \sigma/\sqrt{n}$ to be true, n/N must be less than or equal to .05.

Figure 7.2*a* shows the probability distribution curve for a population. The distribution curves in Figure 7.2*b* through Figure 7.2*e* show the sampling distributions of $\bar{x}$ for different sample sizes taken from the population of Figure 7.2*a*. As we can observe, the population has a normal distribution. Because of this, the sampling distribution of $\bar{x}$ is normal for each of

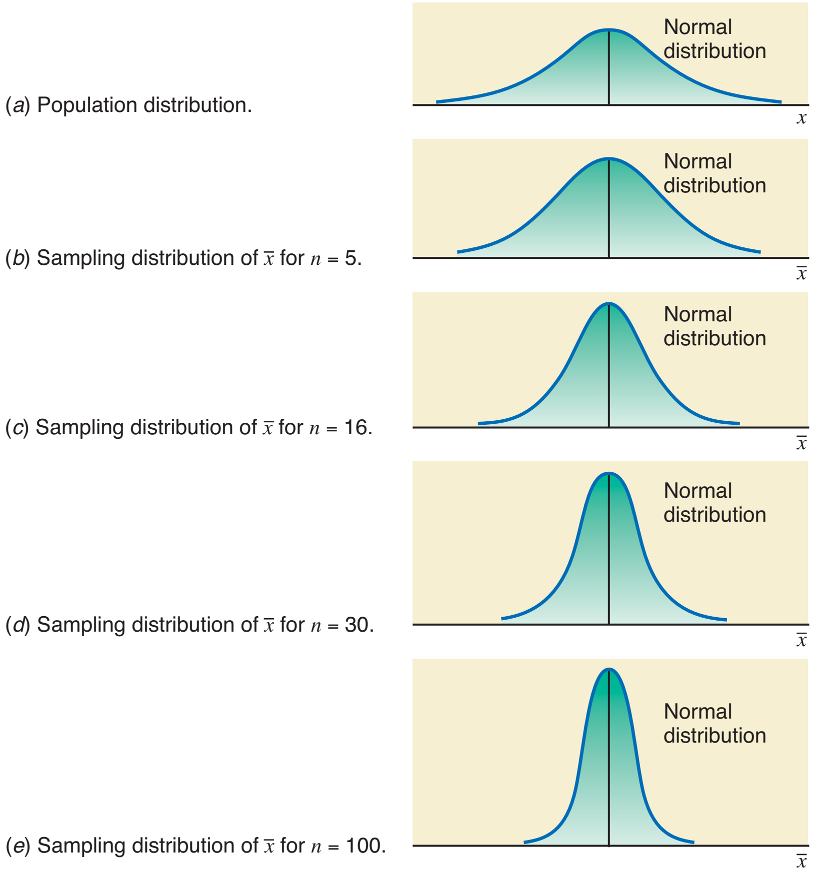


Figure 7.2 Population distribution and sampling distributions of $\bar{x}$.

the four cases illustrated in Figure 7.2*b* through Figure 7.2*e*. Also notice from Figure 7.2*b* through Figure 7.2*e* that the spread of the sampling distribution of $\bar{x}$ decreases as the sample size increases.

Example 7–3 illustrates the calculation of the mean and standard deviation of $\bar{x}$ and the description of the shape of its sampling distribution.

■ EXAMPLE 7–3

Finding the mean, standard deviation, and sampling distribution of $\bar{x}$: normal population.

In a recent SAT, the mean score for all examinees was 1020. Assume that the distribution of SAT scores of all examinees is normal with a mean of 1020 and a standard deviation of 153. Let $\bar{x}$ be the mean SAT score of a random sample of certain examinees. Calculate the mean and standard deviation of $\bar{x}$ and describe the shape of its sampling distribution when the sample size is

(a) 16 **(b)** 50 **(c)** 1000

Solution Let μ and σ be the mean and standard deviation of SAT scores of all examinees, and let $\mu_{\bar{x}}$ and $\sigma_{\bar{x}}$ be the mean and standard deviation of the sampling distribution of $\bar{x}$, respectively. Then, from the given information,

$$\mu = 1020 \quad \text{and} \quad \sigma = 153$$

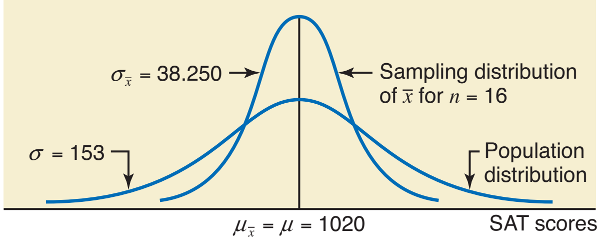
(a) The mean and standard deviation of $\bar{x}$ are, respectively,

$$\mu_{\bar{x}} = \mu = \mathbf{1020} \quad \text{and} \quad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{153}{\sqrt{16}} = \mathbf{38.250}$$

Because the SAT scores of all examinees are assumed to be normally distributed, the sampling distribution of $\bar{x}$ for samples of 16 examinees is also normal. Figure 7.3

shows the population distribution and the sampling distribution of $\bar{x}$. Note that because σ is greater than $\sigma_{\bar{x}}$, the population distribution has a wider spread but smaller height than the sampling distribution of $\bar{x}$ in Figure 7.3.

Figure 7.3

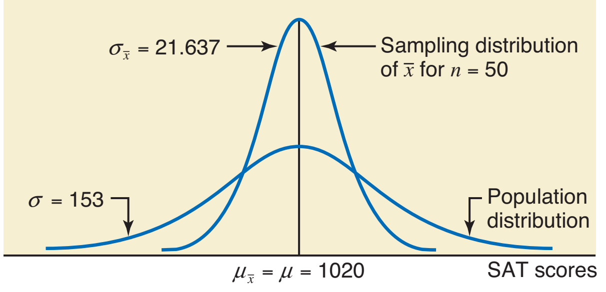


(b) The mean and standard deviation of $\bar{x}$ are, respectively,

$$\mu_{\bar{x}} = \mu = \mathbf{1020} \quad \text{and} \quad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{153}{\sqrt{50}} = \mathbf{21.637}$$

Again, because the SAT scores of all examinees are assumed to be normally distributed, the sampling distribution of $\bar{x}$ for samples of 50 examinees is also normal. The population distribution and the sampling distribution of $\bar{x}$ are shown in Figure 7.4.

Figure 7.4

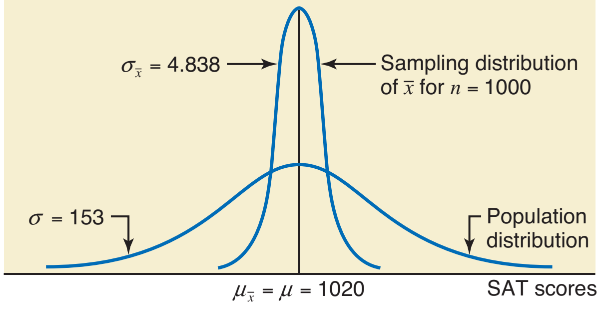


(c) The mean and standard deviation of $\bar{x}$ are, respectively,

$$\mu_{\bar{x}} = \mu = \mathbf{1020} \quad \text{and} \quad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{153}{\sqrt{1000}} = \mathbf{4.838}$$

Again, because the SAT scores of all examinees are assumed to be normally distributed, the sampling distribution of $\bar{x}$ for samples of 1000 examinees is also normal. The two distributions are shown in Figure 7.5.

Figure 7.5



Thus, whatever the sample size, the sampling distribution of $\bar{x}$ is normal when the population from which the samples are drawn is normally distributed. ■

7.3.2 Sampling from a Population That Is Not Normally Distributed

Most of the time the population from which the samples are selected is not normally distributed. In such cases, the shape of the sampling distribution of $\bar{x}$ is inferred from a very important theorem called the **central limit theorem**.

Central Limit Theorem According to the *central limit theorem*, for a large sample size, the sampling distribution of $\bar{x}$ is approximately normal, irrespective of the shape of the population distribution. The mean and standard deviation of the sampling distribution of $\bar{x}$ are, respectively,

$$\mu_{\bar{x}} = \mu \quad \text{and} \quad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

The sample size is usually considered to be large if $n \geq 30$.

Note that when the population does not have a normal distribution, the shape of the sampling distribution is not exactly normal, but it is approximately normal for a large sample size. The approximation becomes more accurate as the sample size increases. Another point to remember is that the central limit theorem applies to *large* samples only. Usually, if the sample size is 30 or larger, it is considered sufficiently large so that the central limit theorem can be applied to the sampling distribution of $\bar{x}$. Thus:

1. When $n \geq 30$, the shape of the sampling distribution of $\bar{x}$ is approximately normal irrespective of the shape of the population distribution. This is so due to the central limit theorem.
2. The mean of $\bar{x}$, $\mu_{\bar{x}}$, is equal to the mean of the population, μ.
3. The standard deviation of $\bar{x}$, $\sigma_{\bar{x}}$, is equal to $\sigma/\sqrt{n}$ if $n/N \leq .05$.

Again, remember that for $\sigma_{\bar{x}} = \sigma/\sqrt{n}$ to apply, n/N must be less than or equal to .05, otherwise we multiply $\sigma/\sqrt{n}$ by the finite population correction factor explained earlier in this chapter.

Figure 7.6*a* shows the probability distribution curve for a population. The distribution curves in Figure 7.6*b* through Figure 7.6*e* show the sampling distributions of $\bar{x}$ for different sample

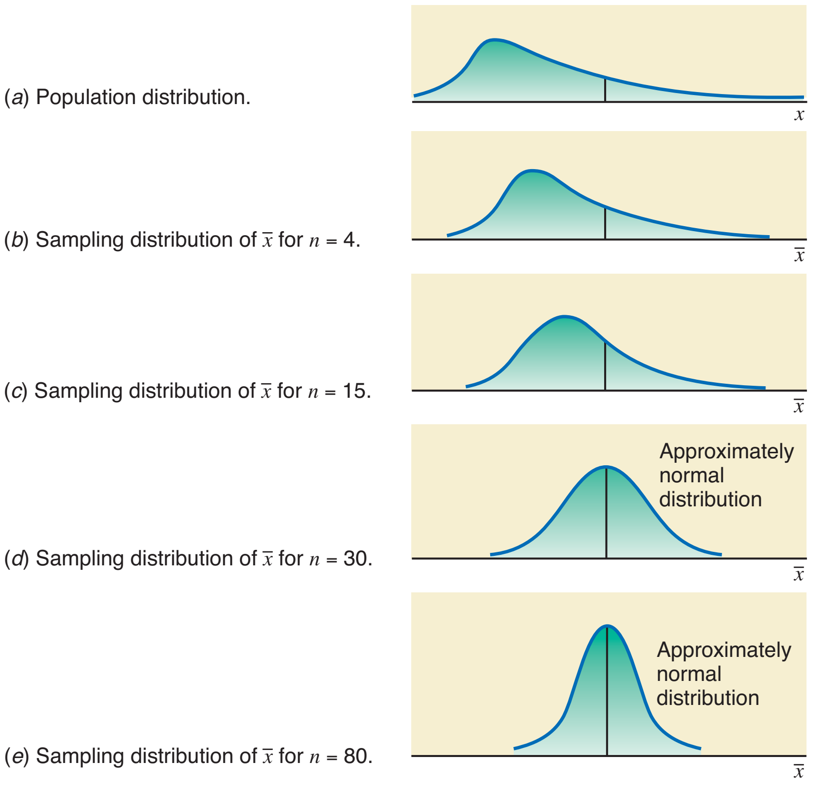


Figure 7.6 Population distribution and sampling distributions of $\bar{x}$.

sizes taken from the population of Figure 7.6*a*. As we can observe, the population is not normally distributed. The sampling distributions of $\bar{x}$ shown in parts *b* and *c*, when $n < 30$, are not normal. However, the sampling distributions of $\bar{x}$ shown in parts *d* and *e*, when $n \geq 30$, are (approximately) normal. Also notice that the spread of the sampling distribution of $\bar{x}$ decreases as the sample size increases.

Example 7–4 illustrates the calculation of the mean and standard deviation of $\bar{x}$ and describes the shape of the sampling distribution of $\bar{x}$ when the sample size is large.

■ EXAMPLE 7–4

Finding the mean, standard deviation, and sampling distribution of $\bar{x}$: nonnormal population.

The mean rent paid by all tenants in a small city is \$1550 with a standard deviation of \$225. However, the population distribution of rents for all tenants in this city is skewed to the right. Calculate the mean and standard deviation of $\bar{x}$ and describe the shape of its sampling distribution when the sample size is

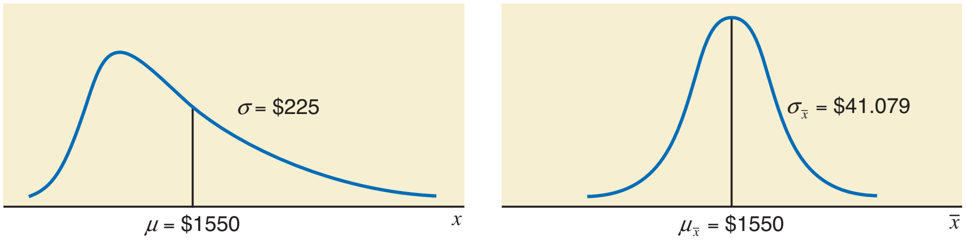
(a) 30 **(b)** 100

Solution Although the population distribution of rents paid by all tenants is not normal, in each case the sample size is large ($n \geq 30$). Hence, the central limit theorem can be applied to infer the shape of the sampling distribution of $\bar{x}$.

(a) Let $\bar{x}$ be the mean rent paid by a sample of 30 tenants. Then, the sampling distribution of $\bar{x}$ is approximately normal with the values of the mean and standard deviation given as

$$\mu_{\bar{x}} = \mu = \mathbf{\$1550} \quad \text{and} \quad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{225}{\sqrt{30}} = \mathbf{\$41.079}$$

Figure 7.7 shows the population distribution and the sampling distribution of $\bar{x}$.



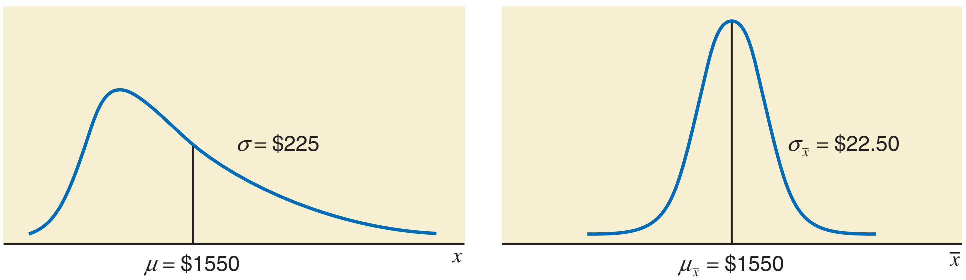
(*a*) Population distribution. (*b*) Sampling distribution of $\bar{x}$ for $n = 30$.

Figure 7.7

(b) Let $\bar{x}$ be the mean rent paid by a sample of 100 tenants. Then, the sampling distribution of $\bar{x}$ is approximately normal with the values of the mean and standard deviation given as

$$\mu_{\bar{x}} = \mu = \mathbf{\$1550} \quad \text{and} \quad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{225}{\sqrt{100}} = \mathbf{\$22.50}$$

Figure 7.8 shows the population distribution and the sampling distribution of $\bar{x}$.



(*a*) Population distribution. (*b*) Sampling distribution of $\bar{x}$ for $n = 100$.

Figure 7.8 ■

EXERCISES

CONCEPTS AND PROCEDURES

7.26 What condition or conditions must hold true for the sampling distribution of the sample mean to be normal when the sample size is less than 30?

7.27 Explain the central limit theorem.

7.28 A population has a distribution that is skewed to the left. Indicate in which of the following cases the central limit theorem will apply to describe the sampling distribution of the sample mean.

a. $n = 500$ **b.** $n = 22$ **c.** $n = 34$

7.29 A population has a distribution that is skewed to the right. A sample of size n is selected from this population. Describe the shape of the sampling distribution of the sample mean for each of the following cases.

a. $n = 25$ **b.** $n = 90$ **c.** $n = 29$

7.30 A population has a normal distribution. A sample of size n is selected from this population. Describe the shape of the sampling distribution of the sample mean for each of the following cases.

a. $n = 98$ **b.** $n = 9$

7.31 A population has a normal distribution. A sample of size n is selected from this population. Describe the shape of the sampling distribution of the sample mean for each of the following cases.

a. $n = 23$ **b.** $n = 450$

APPLICATIONS

7.32 The delivery times for all food orders at a fast-food restaurant during the lunch hour are normally distributed with a mean of 8.4 minutes and a standard deviation of 1.8 minutes. Let $\bar{x}$ be the mean delivery time for a random sample of 16 orders at this restaurant. Calculate the mean and standard deviation of $\bar{x}$, and describe the shape of its sampling distribution.

7.33 Among college students who hold part-time jobs during the school year, the distribution of the time spent working per week is approximately normally distributed with a mean of 20.20 hours and a standard deviation of 2.60 hours. Let $\bar{x}$ be the average time spent working per week for 18 randomly selected college students who hold part-time jobs during the school year. Calculate the mean and the standard deviation of the sampling distribution of $\bar{x}$, and describe the shape of this sampling distribution.

7.34 The amounts of electricity bills for all households in a particular city have an approximately normal distribution with a mean of \$120 and a standard deviation of \$25. Let $\bar{x}$ be the mean amount of electricity bills for a random sample of 25 households selected from this city. Find the mean and standard deviation of $\bar{x}$, and comment on the shape of its sampling distribution.

7.35 The GPAs of all 5540 students enrolled at a university have an approximately normal distribution with a mean of 3.13 and a standard deviation of .52. Let $\bar{x}$ be the mean GPA of a random sample of 48 students selected from this university. Find the mean and standard deviation of $\bar{x}$, and comment on the shape of its sampling distribution.

7.36 The weights of all people living in a particular town have a distribution that is skewed to the right with a mean of 142 pounds and a standard deviation of 31 pounds. Let $\bar{x}$ be the mean weight of a random sample of 45 persons selected from this town. Find the mean and standard deviation of $\bar{x}$ and comment on the shape of its sampling distribution.

7.37 According to an estimate, the average age at first marriage for men in the United States was 28.2 years in 2010 (*Time*, March 21, 2011). Suppose that currently the mean age for all U.S. men at the time of first marriage is 28.2 years with a standard deviation of 6 years and that this distribution is strongly skewed to the right. Let $\bar{x}$ be the average age at the time of first marriage for 25 randomly selected U.S. men. Find the mean and the standard deviation of the sampling distribution of $\bar{x}$. What if the sample size is 100? How do the shapes of the sampling distributions differ for the two sample sizes?

7.38 Suppose that the incomes of all people in the United States who own hybrid (gas and electric) automobiles are normally distributed with a mean of \$78,000 and a standard deviation of \$9100. Let $\bar{x}$ be the mean income of a random sample of 50 owners of such automobiles. Calculate the mean and standard deviation of $\bar{x}$ and describe the shape of its sampling distribution.

7.39 According to the American Time Use Survey, Americans watch television on weekdays for an average of 151 minutes per day (*Time*, July 11, 2011). Suppose that the current distribution of times spent watching television per weekday by all Americans has a mean of 151 minutes and a standard deviation of 20 minutes. Let $\bar{x}$ be the average time spent watching television on a weekday by 200 randomly selected Americans. Find the mean and the standard deviation of the sampling distribution of $\bar{x}$. What is the shape of the sampling distribution of $\bar{x}$? Do you need to know the shape of the population distribution in order to make this conclusion? Explain why or why not.

7.4 Applications of the Sampling Distribution of $\bar{x}$

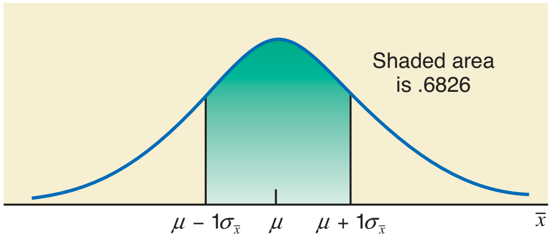
From the central limit theorem, for large samples, the sampling distribution of $\bar{x}$ is approximately normal with mean μ and standard deviation $\sigma_{\bar{x}} = \sigma/\sqrt{n}$. Based on this result, we can make the following statements about $\bar{x}$ for large samples. The areas under the curve of $\bar{x}$ mentioned in these statements are found from the normal distribution table.

1. *If we take all possible samples of the same (large) size from a population and calculate the mean for each of these samples, then about 68.26% of the sample means will be within one standard deviation ($\sigma_{\bar{x}}$) of the population mean.* Alternatively, we can state that if we take one sample (of $n \geq 30$) from a population and calculate the mean for this sample, the probability that this sample mean will be within one standard deviation ($\sigma_{\bar{x}}$) of the population mean is .6826. That is,

$$P(\mu - 1\sigma_{\bar{x}} \leq \bar{x} \leq \mu + 1\sigma_{\bar{x}}) = .8413 - .1587 = .6826$$

This probability is shown in Figure 7.9.

Figure 7.9 $P(\mu - 1\sigma_{\bar{x}} \leq \bar{x} \leq \mu + 1\sigma_{\bar{x}})$

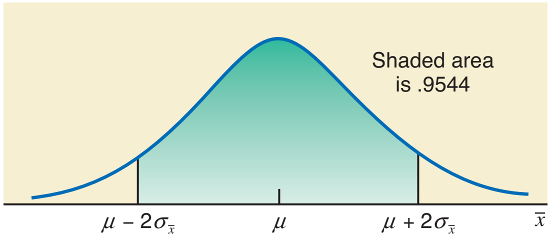


2. *If we take all possible samples of the same (large) size from a population and calculate the mean for each of these samples, then about 95.44% of the sample means will be within two standard deviations ($\sigma_{\bar{x}}$) of the population mean.* Alternatively, we can state that if we take one sample (of $n \geq 30$) from a population and calculate the mean for this sample, the probability that this sample mean will be within two standard deviations ($\sigma_{\bar{x}}$) of the population mean is .9544. That is,

$$P(\mu - 2\sigma_{\bar{x}} \leq \bar{x} \leq \mu + 2\sigma_{\bar{x}}) = .9772 - .0228 = .9544$$

This probability is shown in Figure 7.10.

Figure 7.10 $P(\mu - 2\sigma_{\bar{x}} \leq \bar{x} \leq \mu + 2\sigma_{\bar{x}})$.



3. *If we take all possible samples of the same (large) size from a population and calculate the mean for each of these samples, then about 99.74% of the sample means will be within* *three standard deviations* ($\sigma_{\bar{x}}$) *of the population mean.* Alternatively, we can state that if we take one sample (of $n \geq 30$) from a population and calculate the mean for this sample, the probability that this sample mean will be within three standard deviations ($\sigma_{\bar{x}}$) of the population mean is .9974. That is,

$$P(\mu - 3\sigma_{\bar{x}} \leq \bar{x} \leq \mu + 3\sigma_{\bar{x}}) = .9987 - .0013 = .9974$$

This probability is shown in Figure 7.11.

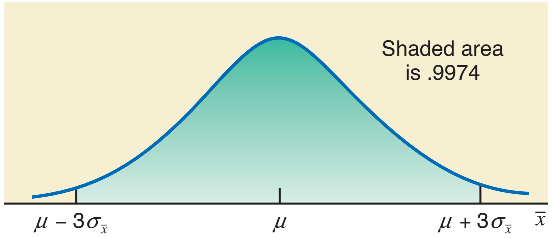


Figure 7.11 $P(\mu - 3\sigma_{\bar{x}} \leq \bar{x} \leq \mu + 3\sigma_{\bar{x}})$

When conducting a survey, we usually select one sample and compute the value of $\bar{x}$ based on that sample. We never select all possible samples of the same size and then prepare the sampling distribution of $\bar{x}$. Rather, we are more interested in finding the probability that the value of $\bar{x}$ computed from one sample falls within a given interval. Examples 7–5 and 7–6 illustrate this procedure.

■ EXAMPLE 7–5

Calculating the probability of $\bar{x}$ in an interval: normal population.

Assume that the weights of all packages of a certain brand of cookies are normally distributed with a mean of 32 ounces and a standard deviation of .3 ounce. Find the probability that the mean weight, $\bar{x}$, of a random sample of 20 packages of this brand of cookies will be between 31.8 and 31.9 ounces.

© Burwell and BurwellPhotography/iStockphoto

Solution Although the sample size is small ($n < 30$), the shape of the sampling distribution of $\bar{x}$ is normal because the population is normally distributed. The mean and standard deviation of $\bar{x}$ are, respectively,

$$\mu_{\bar{x}} = \mu = 32 \text{ ounces} \quad \text{and} \quad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{.3}{\sqrt{20}} = .06708204 \text{ ounce}$$

We are to compute the probability that the value of $\bar{x}$ calculated for one randomly drawn sample of 20 packages is between 31.8 and 31.9 ounces; that is,

$$P(31.8 < \bar{x} < 31.9)$$

This probability is given by the area under the normal distribution curve for $\bar{x}$ between the points $\bar{x} = 31.8$ and $\bar{x} = 31.9$. The first step in finding this area is to convert the two $\bar{x}$ values to their respective z values.

z Value for a Value of $\bar{x}$ The *z value for a value of $\bar{x}$* is calculated as

$$z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}}$$

The z values for $\bar{x} = 31.8$ and $\bar{x} = 31.9$ are computed below, and they are shown on the z scale below the normal distribution curve for $\bar{x}$ in Figure 7.12.

$$\text{For } \bar{x} = 31.8: \quad z = \frac{31.8 - 32}{.06708204} = -2.98$$

$$\text{For } \bar{x} = 31.9: \quad z = \frac{31.9 - 32}{.06708204} = -1.49$$

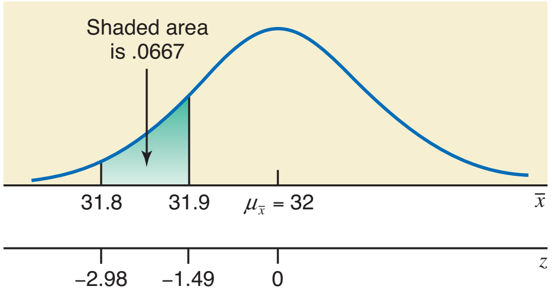


Figure 7.12 $P(31.8 < \bar{x} < 31.9)$

The probability that $\bar{x}$ is between 31.8 and 31.9 is given by the area under the standard normal curve between $z = -2.98$ and $z = -1.49$. Thus, the required probability is

$$P(31.8 < \bar{x} < 31.9) = P(-2.98 < z < -1.49)$$
$$= P(z < -1.49) - P(z < -2.98)$$
$$= .0681 - .0014 = \mathbf{.0667}$$

Therefore, the probability is .0667 that the mean weight of a sample of 20 packages will be between 31.8 and 31.9 ounces. ■

■ EXAMPLE 7–6

Calculating the probability of $\bar{x}$ in an interval: $n \geq 30$.

According to Moebs Services Inc., an individual checking account at major U.S. banks costs the banks between \$350 and \$450 per year (*Time*, November 21, 2011). Suppose that the current average cost of all checking accounts at major U.S. banks is \$400 per year with a standard deviation of \$30. Let $\bar{x}$ be the current average annual cost of a random sample of 225 individual checking accounts at major banks in America.

(a) What is the probability that the average annual cost of the checking accounts in this sample is within \$4 of the population mean?

(b) What is the probability that the average annual cost of the checking accounts in this sample is less than the population mean by \$2.70 or more?

Solution From the given information, for the annual costs of all individual checking accounts at major banks in America,

$$\mu = \$400 \quad \text{and} \quad \sigma = \$30$$

Although the shape of the probability distribution of the population (annual costs of all individual checking accounts at major U.S. banks) is unknown, the sampling distribution of $\bar{x}$ is approximately normal because the sample is large ($n \geq 30$). Remember that when the sample is large, the central limit theorem applies. The mean and standard deviation of the sampling distribution of $\bar{x}$ are, respectively,

$$\mu_{\bar{x}} = \mu = \$400 \quad \text{and} \quad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{30}{\sqrt{225}} = \$2.00$$

(a) The probability that the mean of the annual costs of checking accounts in this sample is within \$4 of the population mean is written as

$$P(396 \leq \bar{x} \leq 404)$$

This probability is given by the area under the normal curve for $\bar{x}$ between $\bar{x} = \$396$ and $\bar{x} = \$404$, as shown in Figure 7.13. We find this area as follows:

For $\bar{x} = \$396$: $\quad z = \dfrac{\bar{x} - \mu}{\sigma_{\bar{x}}} = \dfrac{396 - 400}{2.00} = -2.00$

For $\bar{x} = \$404$: $\quad z = \dfrac{\bar{x} - \mu}{\sigma_{\bar{x}}} = \dfrac{404 - 400}{2.00} = 2.00$

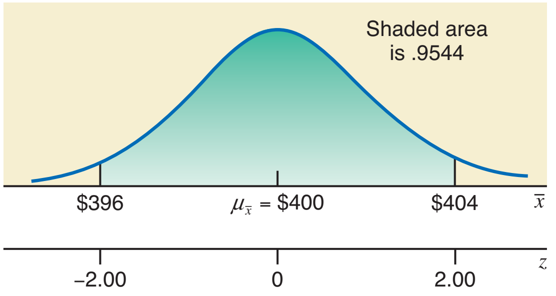


Figure 7.13 $P(\$396 \le \bar{x} \le \$400)$

Hence, the required probability is

$$P(\$396 \le \bar{x} \le \$404) = P(-2.00 \le z \le 2.00)$$
$$= P(z \le 2.00) - P(z \le -2.00)$$
$$= .9772 - .0228 = \mathbf{.9544}$$

Therefore, the probability that the average annual cost of the 225 checking accounts in this sample is within \$4 of the population mean is .9544.

(b) The probability that the average annual cost of the checking accounts in this sample is less than the population mean by \$2.70 or more is written as

$$P(\bar{x} \le 397.30)$$

This probability is given by the area under the normal curve for $\bar{x}$ to the left of $\bar{x} = \$397.30$, as shown in Figure 7.14. We find this area as follows:

For $\bar{x} = \$397.30$: $\quad z = \dfrac{\bar{x} - \mu}{\sigma_{\bar{x}}} = \dfrac{397.30 - 400}{2.00} = -1.35$

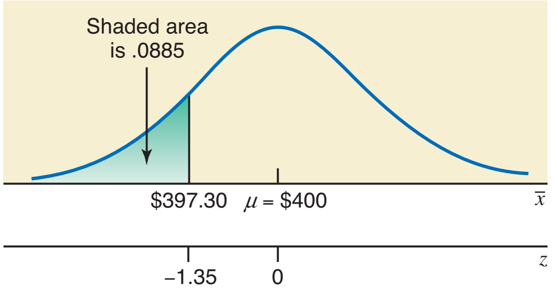


Figure 7.14 $P(\bar{x} \le \$397.30)$

Hence, the required probability is

$$P(\bar{x} \le 397.30) = P(z \le -1.35) = \mathbf{.0885}$$

Thus, the probability that the average annual cost of the checking accounts in this sample is less than the population mean by \$2.70 or more is .0885. ■

EXERCISES

CONCEPTS AND PROCEDURES

7.40 If all possible samples of the same (large) size are selected from a population, what percentage of all the sample means will be within 2.5 standard deviations $(\sigma_{\bar{x}})$ of the population mean?

7.41 If all possible samples of the same (large) size are selected from a population, what percentage of all the sample means will be within 1.5 standard deviations $(\sigma_{\bar{x}})$ of the population mean?

7.42 For a population, $N = 6400$, $\mu = 84$, and $\sigma = 6$. Find the z value for each of the following for $n = 36$.

a. $\bar{x} = 85.60$ **b.** $\bar{x} = 80.20$ **c.** $\bar{x} = 76.64$ **d.** $\bar{x} = 90.05$

7.43 For a population, $N = 205{,}000$, $\mu = 66$, and $\sigma = 14$. Find the z value for each of the following for $n = 49$.

a. $\bar{x} = 66.44$ **b.** $\bar{x} = 58.75$ **c.** $\bar{x} = 62.35$ **d.** $\bar{x} = 72.34$

7.44 Let x be a continuous random variable that has a normal distribution with $\mu = 75$ and $\sigma = 12$. Assuming $n/N \le .05$, find the probability that the sample mean, $\bar{x}$, for a random sample of 20 taken from this population will be

a. between 68.5 and 77.3 **b.** less than 72.4

7.45 Let x be a continuous random variable that has a normal distribution with $\mu = 48$ and $\sigma = 8$. Assuming $n/N \le .05$, find the probability that the sample mean, $\bar{x}$, for a random sample of 16 taken from this population will be

a. between 49.6 and 52.2 **b.** more than 45.7

7.46 Let x be a continuous random variable that has a distribution skewed to the right with $\mu = 60$ and $\sigma = 10$. Assuming $n/N \le .05$, find the probability that the sample mean, $\bar{x}$, for a random sample of 40 taken from this population will be

a. less than 62.40 **b.** between 62.2 and 64.2

7.47 Let x be a continuous random variable that follows a distribution skewed to the left with $\mu = 90$ and $\sigma = 18$. Assuming $n/N \le .05$, find the probability that the sample mean, $\bar{x}$, for a random sample of 64 taken from this population will be

a. less than 82.3 **b.** greater than 86.7

APPLICATIONS

7.48 According to Moebs Services Inc., an individual checking account at U.S. community banks costs these banks between \$175 and \$200 per year (*Time*, November 21, 2011). Suppose that the average annual cost of all such checking accounts at U.S. community banks is \$190 with a standard deviation of \$20. Find the probability that the average annual cost of a random sample of 100 such checking accounts at U.S. community banks is

a. less than \$179 **b.** more than \$191.5 **c.** \$191.70 to 194.5

7.49 The GPAs of all students enrolled at a large university have an approximately normal distribution with a mean of 3.13 and a standard deviation of .52. Find the probability that the mean GPA of a random sample of 20 students selected from this university is

a. 3.20 or higher **b.** 3.00 or lower **c.** 3.05 to 3.21

7.50 The delivery times for all food orders at a fast-food restaurant during the lunch hour are normally distributed with a mean of 8.4 minutes and a standard deviation of 1.8 minutes. Find the probability that the mean delivery time for a random sample of 16 such orders at this restaurant is

a. between 8 and 9 minutes
b. within 1 minute of the population mean
c. less than the population mean by 1 minute or more

7.51 As mentioned in Exercise 7.22, according to the American Automobile Association's 2012 annual report *Your Driving Costs*, the cost of owning and operating a four-wheel drive SUV is \$11,350 per year (*USA TODAY*, April 27, 2012). Note that this cost includes expenses for gasoline, maintenance, insurance, and financing for a vehicle that is driven 15,000 miles a year. Suppose that the distribution of such costs of owning and operating all four-wheel drive SUVs has a mean of \$11,350 with a standard deviation of \$2390. Find the probability that for a random sample of 400 four-wheel drive SUVs, the average cost of owning and operating is

a. more than \$11,540 **b.** less than \$11,110 **c.** \$11,250 to \$11,600

7.52 The times that college students spend studying per week have a distribution that is skewed to the right with a mean of 8.4 hours and a standard deviation of 2.7 hours. Find the probability that the mean time spent studying per week for a random sample of 45 students would be

a. between 8 and 9 hours **b.** less than 8 hours

7.53 The credit card debts of all college students have a distribution that is skewed to the right with a mean of \$2840 and a standard deviation of \$672. Find the probability that the mean credit card debt for a random sample of 36 college students would be

a. between \$2600 and \$2950 **b.** less than \$3060

7.54 As mentioned in Exercise 7.39, according to the American Time Use Survey, Americans watch television each weekday for an average of 151 minutes (*Time*, July 11, 2011). Suppose that the current distribution of times spent watching television every weekday by all Americans has a mean of 151 minutes and a standard deviation of 20 minutes. Find the probability that the average time spent watching television on a weekday by 200 randomly selected Americans is

a. 148.70 to 150 minutes **b.** more than 153 minutes **c.** at most 146 minutes

7.55 The amounts of electricity bills for all households in a city have a skewed probability distribution with a mean of \$140 and a standard deviation of \$30. Find the probability that the mean amount of electric bills for a random sample of 75 households selected from this city will be

a. between \$132 and \$136
b. within \$6 of the population mean
c. more than the population mean by at least \$4

7.56 According to a PNC Financial Independence Survey released in March 2012, today's U.S. adults in their 20s "hold an average debt of about \$45,000, which includes everything from cars to credit cards to student loans to mortgages" (*USA TODAY*, April 24, 2012). Suppose that the current distribution of debts of all U.S. adults in their 20s has a mean of \$45,000 and a standard deviation of \$12,720. Find the probability that the average debt of a random sample of 144 U.S. adults in their 20s is

a. less than \$42,600 **b.** more than \$46,240 **c.** \$43,190 to \$46,980

7.57 As mentioned in Exercise 7.33, among college students who hold part-time jobs during the school year, the distribution of the time spent working per week is approximately normally distributed with a mean of 20.20 hours and a standard deviation of 2.60 hours. Find the probability that the average time spent working per week for 18 randomly selected college students who hold part-time jobs during the school year is

a. not within 1 hour of the population mean
b. 20 to 20.50 hours
c. at least 22 hours
d. no more than 21 hours

7.58 Johnson Electronics Corporation makes electric tubes. It is known that the standard deviation of the lives of these tubes is 150 hours. The company's research department takes a sample of 100 such tubes and finds that the mean life of these tubes is 2250 hours. What is the probability that this sample mean is within 25 hours of the mean life of all tubes produced by this company?

7.59 A machine at Katz Steel Corporation makes 3-inch-long nails. The probability distribution of the lengths of these nails is normal with a mean of 3 inches and a standard deviation of .1 inch. The quality control inspector takes a sample of 25 nails once a week and calculates the mean length of these nails. If the mean of this sample is either less than 2.95 inches or greater than 3.05 inches, the inspector concludes that the machine needs an adjustment. What is the probability that based on a sample of 25 nails, the inspector will conclude that the machine needs an adjustment?

7.5 Population and Sample Proportions; and Mean, Standard Deviation, and Shape of the Sampling Distribution of $\hat{p}$

The concept of proportion is the same as the concept of relative frequency discussed in Chapter 2 and the concept of probability of success in a binomial experiment. The relative frequency of a category or class gives the proportion of the sample or population that belongs to that category or class. Similarly, the probability of success in a binomial experiment represents the proportion of the sample or population that possesses a given characteristic.

In this section, first we will learn about the population and sample proportions. Then we will discuss the mean, standard deviation and shape of the sampling distribution of $\hat{p}$.

7.5.1 Population and Sample Proportions

The **population proportion**, denoted by $\boldsymbol{p}$, is obtained by taking the ratio of the number of elements in a population with a specific characteristic to the total number of elements in the population. The **sample proportion**, denoted by $\hat{\boldsymbol{p}}$ (pronounced *p hat*), gives a similar ratio for a sample.

Population and Sample Proportions The *population* and *sample proportions*, denoted by p and $\hat{p}$, respectively, are calculated as

$$p = \frac{X}{N} \quad \text{and} \quad \hat{p} = \frac{x}{n}$$

where

N = total number of elements in the population
n = total number of elements in the sample
X = number of elements in the population that possess a specific characteristic
x = number of elements in the sample that possess a specific characteristic

Example 7–7 illustrates the calculation of the population and sample proportions.

■ EXAMPLE 7–7

Calculating the population and sample proportions.

Suppose a total of 789,654 families live in a particular city and 563,282 of them own homes. A sample of 240 families is selected from this city, and 158 of them own homes. Find the proportion of families who own homes in the population and in the sample.

Solution For the population of this city,

$$N = \text{population size} = 789{,}654$$

$$X = \text{families in the population who own homes} = 563{,}282$$

The proportion of all families in this city who own homes is

$$p = \frac{X}{N} = \frac{563{,}282}{789{,}654} = \mathbf{.71}$$

Now, a sample of 240 families is taken from this city, and 158 of them are home-owners. Then,

$$n = \text{sample size} = 240$$

$$x = \text{families in the sample who own homes} = 158$$

The sample proportion is

$$\hat{p} = \frac{x}{n} = \frac{158}{240} = \mathbf{.66}$$ ■

As in the case of the mean, the difference between the sample proportion and the corresponding population proportion gives the sampling error, assuming that the sample is random and no nonsampling error has been made. Thus, in the case of the proportion,

$$\text{Sampling error} = \hat{p} - p$$

For instance, for Example 7–7,

$$\text{Sampling error} = \hat{p} - p = .66 - .71 = -.05$$

7.5.2 Sampling Distribution of $\hat{p}$

Just like the sample mean $\bar{x}$, the sample proportion $\hat{p}$ is a random variable. In other words, the population proportion p is a contant as it assumes one and only one value. However, the sample proportion $\hat{p}$ can assume one of a large number of possible values depending on which sample is selected. Hence, $\hat{p}$ is a random variable and it possesses a probability distribution, which is called its **sampling distribution**.

Definition

Sampling Distribution of the Sample Proportion, $\hat{p}$ The probability distribution of the sample proportion, $\hat{p}$, is called its *sampling distribution*. It gives the various values that $\hat{p}$ can assume and their probabilities.

The value of $\hat{p}$ calculated for a particular sample depends on what elements of the population are included in that sample. Example 7–8 illustrates the concept of the sampling distribution of $\hat{p}$.

■ EXAMPLE 7–8

Boe Consultant Associates has five employees. Table 7.6 gives the names of these five employees and information concerning their knowledge of statistics.

Illustrating the sampling distribution of $\hat{p}$.

Table 7.6 **Information on the Five Employees of Boe Consultant Associates**

Name	Knows Statistics
Ally	Yes
John	No
Susan	No
Lee	Yes
Tom	Yes

If we define the population proportion, p, as the proportion of employees who know statistics, then

$$p = 3/5 = .60$$

Now, suppose we draw all possible samples of three employees each and compute the proportion of employees, for each sample, who know statistics. The total number of samples of size three that can be drawn from the population of five employees is

$$\text{Total number of samples} = {}_5C_3 = \frac{5!}{3!(5-3)!} = \frac{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{3 \cdot 2 \cdot 1 \cdot 2 \cdot 1} = 10$$

Table 7.7 lists these 10 possible samples and the proportion of employees who know statistics for each of those samples. Note that we have rounded the values of $\hat{p}$ to two decimal places.

Table 7.7 All Possible Samples of Size 3 and the Value of $\hat{p}$ for Each Sample

Sample	Proportion Who Know Statistics $\hat{p}$
Ally, John, Susan	1/3 = .33
Ally, John, Lee	2/3 = .67
Ally, John, Tom	2/3 = .67
Ally, Susan, Lee	2/3 = .67
Ally, Susan, Tom	2/3 = .67
Ally, Lee, Tom	3/3 = 1.00
John, Susan, Lee	1/3 = .33
John, Susan, Tom	1/3 = .33
John, Lee, Tom	2/3 = .67
Susan, Lee, Tom	2/3 = .67

Using Table 7.7, we prepare the frequency distribution of $\hat{p}$ as recorded in Table 7.8, along with the relative frequencies of classes, which are obtained by dividing the frequencies of classes by the population size. The relative frequencies are used as probabilities and listed in Table 7.9. This table gives the sampling distribution of $\hat{p}$.

Table 7.8 Frequency and Relative Frequency Distributions of $\hat{p}$ When the Sample Size Is 3

$\hat{p}$	f	Relative Frequency
.33	3	3/10 = .30
.67	6	6/10 = .60
1.00	1	1/10 = .10
	$\Sigma f = 10$	Sum = 1.00

Table 7.9 Sampling Distribution of $\hat{p}$ When the Sample Size Is 3

$\hat{p}$	$P(\hat{p})$
.33	.30
.67	.60
1.00	.10
	$\Sigma P(\hat{p}) = 1.00$

■

7.5.3 Mean and Standard Deviation of $\hat{p}$

The **mean of $\hat{p}$**, which is the same as the mean of the sampling distribution of $\hat{p}$, is always equal to the population proportion, p, just as the mean of the sampling distribution of $\bar{x}$ is always equal to the population mean, μ.

Mean of the Sample Proportion The *mean of the sample proportion*, $\hat{p}$, is denoted by $\mu_{\hat{p}}$ and is equal to the population proportion, p. Thus,

$$\mu_{\hat{p}} = p$$

The sample proportion, $\hat{p}$, is called an **estimator** of the population proportion, p. As mentioned earlier, when the expected value (or mean) of a sample statistic is equal to the value of the corresponding population parameter, that sample statistic is said to be an **unbiased estimator**. Since for the sample proportion $\mu_{\hat{p}} = p$, $\hat{p}$ is an unbiased estimator of p.

The **standard deviation of $\hat{p}$**, denoted by $\sigma_{\hat{p}}$, is given by the following formula. This formula is true only when the sample size is small compared to the population size. As we know from Section 7.2, the sample size is said to be small compared to the population size if $n/N \leq .05$.

Standard Deviation of the Sample Proportion The *standard deviation of the sample proportion,* $\hat{p}$, is denoted by $\sigma_{\hat{p}}$ and is given by the formula

$$\sigma_{\hat{p}} = \sqrt{\frac{pq}{n}}$$

where p is the population proportion, $q = 1 - p$, and n is the sample size. This formula is used when $n/N \leq .05$, where N is the population size.

However, if n/N is greater than .05, then $\sigma_{\hat{p}}$ is calculated as follows:

$$\sigma_{\hat{p}} = \sqrt{\frac{pq}{n}}\sqrt{\frac{N-n}{N-1}}$$

where the factor

$$\sqrt{\frac{N-n}{N-1}}$$

is called the finite population correction factor.

In almost all cases, the sample size is small compared to the population size and, consequently, the formula used to calculate $\sigma_{\hat{p}}$ is $\sqrt{pq/n}$.

As mentioned earlier, if the standard deviation of a sample statistic decreases as the sample size is increased, that statistic is said to be a **consistent estimator**. It is obvious from the above formula for $\sigma_{\hat{p}}$ that as n increases, the value of $\sqrt{pq/n}$ decreases. Thus, the sample proportion, $\hat{p}$, is a consistent estimator of the population proportion, p.

7.5.4 Shape of the Sampling Distribution of $\hat{p}$

The shape of the sampling distribution of $\hat{p}$ is inferred from the central limit theorem.

Central Limit Theorem for Sample Proportion According to the central limit theorem, the *sampling distribution of* $\hat{p}$ is approximately normal for a sufficiently large sample size. In the case of proportion, the sample size is considered to be sufficiently large if np and nq are both greater than 5; that is, if

$$np > 5 \quad \text{and} \quad nq > 5$$

Note that the sampling distribution of $\hat{p}$ will be approximately normal if $np > 5$ and $nq > 5$. This is the same condition that was required for the application of the normal approximation to the binomial probability distribution in Chapter 6.

Example 7–9 shows the calculation of the mean and standard deviation of $\hat{p}$ and describes the shape of its sampling distribution.

■ EXAMPLE 7–9

Finding the mean and standard deviation, and describing the shape of the sampling distribution of $\hat{p}$.

According to a *New York Times*/CBS News poll conducted during June 24–28, 2011, 55% of adults polled said that owning a home is a *very* important part of the American Dream (*The New York Times*, June 30, 2011). Assume that this result is true for the current population of American adults. Let $\hat{p}$ be the proportion of American adults in a random sample of 2000 who will say that owning a home is a *very* important part of the American Dream. Find the mean and standard deviation of $\hat{p}$ and describe the shape of its sampling distribution.

Solution Let p be the proportion of all American adults who will say that owning a home is a *very* important part of the American Dream. Then,

$$p = .55, \qquad q = 1 - p = 1 - .55 = .45, \qquad \text{and} \qquad n = 2000$$

The mean of the sampling distribution of $\hat{p}$ is

$$\mu_{\hat{p}} = p = \mathbf{.55}$$

The standard deviation of $\hat{p}$ is

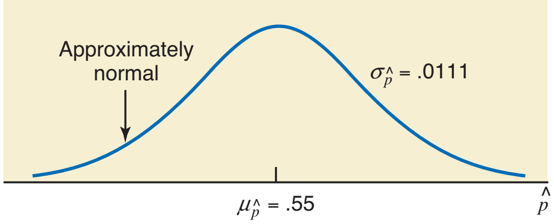
$$\sigma_{\hat{p}} = \sqrt{\frac{pq}{n}} = \sqrt{\frac{(.55)(.45)}{2000}} = \mathbf{.0111}$$

The values of np and nq are

$$np = 2000(.55) = 1100 \qquad \text{and} \qquad nq = 2000(.45) = 900$$

Because np and nq are both greater than 5, we can apply the central limit theorem to make an inference about the shape of the sampling distribution of $\hat{p}$. Thus, the sampling distribution of $\hat{p}$ is approximately normal with a mean of .55 and a standard deviation of .0111, as shown in Figure 7.15.

Figure 7.15 Sampling distribution of $\hat{p}$.



EXERCISES

CONCEPTS AND PROCEDURES

7.60 In a population of 1000 subjects, 580 possess a certain characteristic. A sample of 50 subjects selected from this population has 32 subjects who possess the same characteristic. What are the values of the population and sample proportions?

7.61 In a population of 5000 subjects, 800 possess a certain characteristic. A sample of 150 subjects selected from this population contains 20 subjects who possess the same characteristic. What are the values of the population and sample proportions?

7.62 In a population of 16,400 subjects, 30% possess a certain characteristic. In a sample of 200 subjects selected from this population, 25% possess the same characteristic. How many subjects in the population and sample, respectively, possess this characteristic?

7.63 In a population of 9500 subjects, 75% possess a certain characteristic. In a sample of 400 subjects selected from this population, 78% possess the same characteristic. How many subjects in the population and sample, respectively, possess this characteristic?

7.64 Let $\hat{p}$ be the proportion of elements in a sample that possess a characteristic.

a. What is the mean of $\hat{p}$?

b. What is the formula to calculate the standard deviation of $\hat{p}$? Assume $n/N \leq .05$.

c. What condition(s) must hold true for the sampling distribution of $\hat{p}$ to be approximately normal?

7.65 For a population, $N = 10{,}000$ and $p = .68$. A random sample of 900 elements selected from this population gave $\hat{p} = .64$. Find the sampling error.

7.66 For a population, $N = 2800$ and $p = .28$. A random sample of 100 elements selected from this population gave $\hat{p} = .33$. Find the sampling error.

7.67 What is the estimator of the population proportion? Is this estimator an unbiased estimator of p? Explain why or why not.

7.68 Is the sample proportion a consistent estimator of the population proportion? Explain why or why not.

7.69 How does the value of $\sigma_{\hat{p}}$ change as the sample size increases? Explain. Assume $n/N \leq .05$.

7.70 Consider a large population with $p = .82$. Assuming $n/N \leq .05$, find the mean and standard deviation of the sample proportion $\hat{p}$ for a sample size of

a. 100 **b.** 1000

7.71 Consider a large population with $p = .34$. Assuming $n/N \leq .05$, find the mean and standard deviation of the sample proportion $\hat{p}$ for a sample size of

a. 300 **b.** 850

7.72 A population of $N = 4000$ has a population proportion equal to .12. In each of the following cases, which formula will you use to calculate $\sigma_{\hat{p}}$ and why? Using the appropriate formula, calculate $\sigma_{\hat{p}}$ for each of these cases.

a. $n = 800$ **b.** $n = 30$

7.73 A population of $N = 1400$ has a population proportion equal to .47. In each of the following cases, which formula will you use to calculate $\sigma_{\hat{p}}$ and why? Using the appropriate formula, calculate $\sigma_{\hat{p}}$ for each of these cases.

a. $n = 90$ **b.** $n = 50$

7.74 According to the central limit theorem, the sampling distribution of $\hat{p}$ is approximately normal when the sample is large. What is considered a large sample in the case of the proportion? Briefly explain.

7.75 Indicate in which of the following cases the central limit theorem will apply to describe the sampling distribution of the sample proportion.

a. $n = 400$ and $p = .28$ **b.** $n = 80$ and $p = .05$
c. $n = 60$ and $p = .12$ **d.** $n = 100$ and $p = .035$

7.76 Indicate in which of the following cases the central limit theorem will apply to describe the sampling distribution of the sample proportion.

a. $n = 20$ and $p = .45$ **b.** $n = 75$ and $p = .22$
c. $n = 350$ and $p = .01$ **d.** $n = 200$ and $p = .022$

APPLICATIONS

7.77 A company manufactured five television sets on a given day, and these TV sets were inspected for being good or defective. The results of the inspection follow.

Good Good Defective Defective Good

a. What proportion of these TV sets are good?
b. How many total samples (without replacement) of size four can be selected from this population?
c. List all the possible samples of size four that can be selected from this population and calculate the sample proportion, $\hat{p}$, of television sets that are good for each sample. Prepare the sampling distribution of $\hat{p}$.
d. For each sample listed in part c, calculate the sampling error.

7.78 Investigation of all six major fires in a western desert during one of the recent summers found the following causes.

Arson Accident Accident Arson Accident Accident

a. What proportion of those fires were due to arson?
b. How many total samples (without replacement) of size two can be selected from this population?
c. List all the possible samples of size two that can be selected from this population and calculate the sample proportion $\hat{p}$ of the fires due to arson for each sample. Prepare the table that gives the sampling distribution of $\hat{p}$.
d. For each sample listed in part c, calculate the sampling error.

7.79 Beginning in the second half of 2011, there were widespread protests in many American cities that were primarily against Wall Street corruption and the gap between the rich and the poor in America. According to a *Time* Magazine/ABT SRBI poll conducted by telephone during October 9–10, 2011, 86% of adults who were familiar with those protests agreed that Wall Street and lobbyists have too much influence in Washington (*The New York Times*, October 22, 2011). Assume that this percentage is true for the current population of American adults. Let $\hat{p}$ be the proportion in a random sample of 400 American adults who hold the opinion that Wall Street and lobbyists have too much influence in Washington. Find the mean and standard deviation of the sampling distribution of $\hat{p}$ and describe its shape.

7.80 According to a poll, 55% of Americans do not know that GOP stands for Grand Old Party (*Time*, October 17, 2011). Assume that this percentage is true for the current population of Americans. Let $\hat{p}$ be the proportion in a random sample of 800 Americans who do not know that GOP stands for Grand Old Party. Find the mean and standard deviation of the sampling distribution of $\hat{p}$ and describe its shape.

7.81 In a *Time/Money Magazine* poll of Americans age 18 years and older, 65% agreed with the statement, "We are less sure our children will achieve the American Dream" (*Time*, October 10, 2011). Assume that this result is true for the current population of Americans age 18 years and older. Let $\hat{p}$ be the proportion in a random sample of 600 Americans age 18 years and older who agree with the above statement. Find the mean and standard deviation of the sampling distribution of $\hat{p}$ and describe its shape.

7.82 In a *Time* Magazine/Aspen poll of American adults conducted by the strategic research firm Penn Schoen Berland, these adults were asked, "In your opinion, what is more important for the U.S. to focus on in the next decade?" Eighty-three percent of the adults polled said *domestic issues* (*Time*, July 11, 2011). Assume that this percentage is true for the current population of American adults. Let $\hat{p}$ be the proportion in a random sample of 1000 American adults who hold the above opinion. Find the mean and standard deviation of the sampling distribution of $\hat{p}$ and describe its shape.

7.6 Applications of the Sampling Distribution of $\hat{p}$

As mentioned in Section 7.4, when we conduct a study, we usually take only one sample and make all decisions or inferences on the basis of the results of that one sample. We use the concepts of the mean, standard deviation, and shape of the sampling distribution of $\hat{p}$ to determine the probability that the value of $\hat{p}$ computed from one sample falls within a given interval. Examples 7–10 and 7–11 illustrate this application.

■ EXAMPLE 7–10

Calculating the probability that $\hat{p}$ is in an interval.

According to a Pew Research Center nationwide telephone survey of American adults conducted by phone between March 15 and April 24, 2011, 75% of adults said that college education has become too expensive for most people and they cannot afford it (*Time*, May 30, 2011). Suppose that this result is true for the current population of American adults. Let $\hat{p}$ be the proportion in a random sample of 1400 adult Americans who will hold the said opinion. Find the probability that 76.5% to 78% of adults in this sample will hold this opinion.

Solution From the given information,

$$n = 1400, \qquad p = .75, \qquad \text{and} \qquad q = 1 - p = 1 - .75 = .25$$

where p is the proportion of all adult Americans who hold the said opinion.

The mean of the sample proportion $\hat{p}$ is

$$\mu_{\hat{p}} = p = .75$$

The standard deviation of $\hat{p}$ is

$$\sigma_{\hat{p}} = \sqrt{\frac{pq}{n}} = \sqrt{\frac{(.75)(.25)}{1400}} = .01157275$$

The values of np and nq are

$$np = 1400\,(.75) = 1050 \qquad \text{and} \qquad nq = 1400\,(.25) = 350$$

Because np and nq are both greater than 5, we can infer from the central limit theorem that the sampling distribution of $\hat{p}$ is approximately normal. The probability that $\hat{p}$ is between .765 and .78 is given by the area under the normal curve for $\hat{p}$ between $\hat{p} = .765$ and $\hat{p} = .78$, as shown in Figure 7.16.