PART 4

Inference: From Samples to Population

In Chapter 1, we presented the following process of statistics:

***Step 1*:** Identify a research objective.

***Step 2*:** Collect the information needed to answer the questions posed in Step 1.

***Step 3*:** Organize and summarize the information.

***Step 4*:** Draw conclusions from the information.

The methods for conducting Steps 1 and 2 were discussed in Chapter 1. The methods for conducting Step 3 were discussed in Chapters 2 through 4. We took a break from the statistical process in Chapters 5 through 7 so that we could develop skills that allow us to tackle Step 4.

If the information (data) collected is from a population, we can use the summaries obtained in Step 3 to draw conclusions about the population being studied and the statistical process is over.

However, it is often difficult or impossible to gain access to populations, so the information obtained in Step 2 is often sample data. The sample data are used to make inferences about the population. For example, we might compute a sample mean from the information collected in Step 2 and use this information to draw conclusions regarding the population mean. The last part of this text discusses how sample data are used to draw conclusions about populations.

Sampling Distributions

CHAPTER 8

Outline

DECISIONS

The American Time Use Survey is a survey of adult Americans conducted by the Bureau of Labor Statistics. The purpose of the survey is to learn how Americans allocate their time during a day. As a reporter for the school newspaper, you wish to file a report that compares the typical student at your school to the rest of Americans. See the Decisions project on page 388.

Putting It All Together

In Chapters 6 and 7, we learned about random variables and their probability distributions. A random variable is a numerical measure of the outcome to a probability experiment. A probability distribution provides a way to assign probabilities to the random variable. For discrete random variables, we discussed the binomial probability distribution. We assigned probabilities using a formula. For continuous random variables, we discussed the normal probability distribution. To compute probabilities for a normal random variable, we found the area under a normal density curve.

In this chapter, we continue our discussion of probability distributions where statistics, such as $\bar{x}$, will be the random variable. Statistics are random variables because the value of a statistic varies from sample to sample. For this reason, statistics have probability distributions associated with them. For example, there is a probability distribution for the sample mean, sample variance, and so on. We use probability distributions to make probability statements regarding the statistic. So this chapter discusses the shape, center, and spread of statistics such as $\bar{x}$.

8.1 Distribution of the Sample Mean

Preparing for This Section Before getting started, review the following:

- Simple random sampling (Section 1.2, pp. 16–19)
- The mean (Section 3.1, pp. 107–110)
- The standard deviation (Section 3.2, pp. 129–130)
- Applications of the normal distribution (Section 7.3, pp. 345–349)

Objectives

1. **Understand the concept of a sampling distribution**
2. **Describe the distribution of the sample mean for samples obtained from normal populations**
3. **Describe the distribution of the sample mean for samples obtained from a population that is not normal**

Suppose that the government wanted to estimate the mean income of all U.S. households. One approach the government could take is to literally survey each household in the United States to determine the population mean, μ. This would be a very expensive and time-consuming survey!

A second approach that the government could (and does) take is to survey a random sample of U.S. households and use the results of the survey to estimate the mean household income. This is done through the American Community Survey. The survey is administered to approximately 250,000 randomly selected households each month. Among the many questions on the survey, respondents are asked to report the income of each individual in the household. From this information, the federal government obtains a sample mean household income for U.S. households. For example, in 2003 the mean annual household income in the United States was estimated to be $\bar{x} = \$58{,}036$. The government might infer from this result that the mean annual household income of *all* U.S. households in 2003 was $\mu = \$58{,}036$. This type of statement is an example of **statistical inference**—using information from a sample to draw conclusions about a population.

The households that were administered the American Community Survey were determined by chance (random sampling). A second random sample of households would likely lead to a different sample mean such as $\bar{x} = \$58{,}132$, and a third random sample of households would likely lead to a third distinct sample mean such as $\bar{x} = \$58{,}095$. Because the households are selected by chance, the sample mean of household income is also determined by chance. We conclude from this that there is variability in our estimates. This variability leads to uncertainty as to whether our estimates are correct. Therefore, we need a way to assess the reliability of inferences made about a population based on sample data.

The measure of reliability is actually a statement of probability. Probability describes how likely an outcome is to occur. The goal of this chapter is to learn the distribution of statistics such as the sample mean so that our estimates are accompanied by statements that indicate the likelihood that our methods are accurate.

1 Understand the Concept of a Sampling Distribution

In general, the sampling distribution of a statistic is a probability distribution for all possible values of the statistic computed from a sample of size n. The **sampling distribution of the sample mean** is the probability distribution of all possible values of the random variable $\bar{x}$ computed from a sample of size n from a population with mean μ and standard deviation σ.

In Other Words
If the number of individuals in a population is a positive integer, we say the population is finite. Otherwise, the population is infinite.

The idea behind obtaining the sampling distribution of the mean is as follows:

***Step 1*:** Obtain a simple random sample of size n.

***Step 2*:** Compute the sample mean.

***Step 3*:** Assuming that we are sampling from a finite population, repeat Steps 1 and 2 until all simple random samples of size n have been obtained. **Note:** Once a particular sample is obtained, it cannot be obtained a second time.

We present an example to illustrate the idea behind a sampling distribution.

EXAMPLE 1 A Sampling Distribution

***Problem*:** One semester, Professor Goehl had a small statistics class of seven students. He asked them the ages of their cars and obtained the following data:

$$2, 4, 6, 8, 4, 3, 7$$

Construct a sampling distribution of the mean for samples of size $n = 2$. What is the probability of obtaining a sample mean between 4 and 6 years, inclusive; that is, what is $P(4 \leq \bar{x} \leq 6)$?

***Approach*:** We follow Steps 1 to 3 listed above to construct the probability distribution.

***Solution*:** There are seven individuals in the population. We are selecting them two at a time without replacement. Therefore, there are ${}_7C_2 = 21$ samples of size $n = 2$. We list these 21 samples along with the sample means in Table 1.

Table 1

Sample	Sample Mean	Sample	Sample Mean	Sample	Sample Mean
2, 4	3	4, 8	6	6, 7	6.5
2, 6	4	4, 4	4	8, 4	6
2, 8	5	4, 3	3.5	8, 3	5.5
2, 4	3	4, 7	5.5	8, 7	7.5
2, 3	2.5	6, 8	7	4, 3	3.5
2, 7	4.5	6, 4	5	4, 7	5.5
4, 6	5	6, 3	4.5	3, 7	5

Table 2 displays the sampling distribution of the sample mean, $\bar{x}$.

Table 2

Sample Mean	Frequency	Probability	Sample Mean	Frequency	Probability
2.5	1	$\frac{1}{21}$	5.5	3	$\frac{3}{21}$
3	2	$\frac{2}{21}$	6	2	$\frac{2}{21}$
3.5	2	$\frac{2}{21}$	6.5	1	$\frac{1}{21}$
4	2	$\frac{2}{21}$	7	1	$\frac{1}{21}$
4.5	2	$\frac{2}{21}$	7.5	1	$\frac{1}{21}$
5	4	$\frac{4}{21}$			

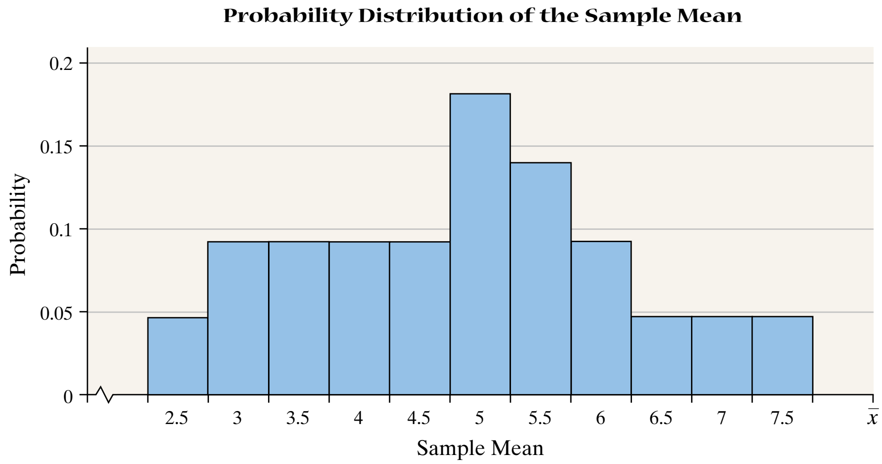
From Table 2 we can compute

$$P(4 \leq \bar{x} \leq 6) = \frac{2}{21} + \frac{2}{21} + \frac{4}{21} + \frac{3}{21} + \frac{2}{21} = \frac{13}{21} = 0.619$$

If we took 10 simple random samples of size 2 from this population, about 6 of them would result in sample means between 4 and 6 years, inclusive.

The sample mean with the highest probability is $\bar{x} = 5$. This should not be surprising since the population mean of the data in Example 1 is $\mu = 4.9$, rounded to one decimal place. Figure 1 is a probability histogram of the sampling distribution for the sample mean given in Table 2.

Figure 1



Now Work Problem 31.

In-Class Activity: Sampling Distributions

Randomly select six students from the class to treat as a population. Choose a quantitative variable (such as pulse rate, age, or number of siblings) to use for this activity, and gather the data appropriately. Compute μ for the population. Divide the class into four groups and have one group list all samples of size $n = 2$, another group list all samples of size $n = 3$, and other groups list all samples of size $n = 4$ and $n = 5$. Each group should do the following:

(a) Compute the sample mean of each sample.
(b) Form the probability distribution for the sample mean.
(c) Draw a probability histogram of the probability distribution.
(d) Verify that $\mu_{\bar{x}} = \mu$.

Compare the spread in each probability distribution based on the probability histogram. What does this result imply about the standard deviation of the sample mean?

2 Describe the Distribution of the Sample Mean for Samples Obtained from Normal Populations

The point of Example 1 is to help you realize that statistics such as $\bar{x}$ are random variables and therefore have probability distributions associated with them. In practice, a single random sample of size n is obtained from a population. The probability distribution of the sample statistic (or sampling distribution) is determined from statistical theory. We will use simulation to help justify the result that statistical theory provides. We consider two possibilities. In the first case, we sample from a population that is known to be normally distributed. In the second case, we sample from a distribution that is not normally distributed.

EXAMPLE 2 Sampling Distribution of the Sample Mean: Population Normal

Problem: In Example 3 from Section 7.1, we learned that the height of 3-year-old females is approximately normally distributed with $\mu = 38.72$ inches and $\sigma = 3.17$ inches. Approximate the sampling distribution of $\overline{x}$ by taking 100 simple random samples of size $n = 5$.

Approach: Use MINITAB, Excel, or some other statistical software package to perform the simulation. We will perform the following steps:

***Step 1*:** Obtain 100 simple random samples of size $n = 5$ from the population, using simulation.

***Step 2*:** Compute the mean of each sample.

***Step 3*:** Draw a histogram of the sample means.

***Step 4*:** Compute the mean and standard deviation of the sample means.

Solution

***Step 1*:** We obtain 100 simple random samples of size $n = 5$. All the samples of size $n = 5$ are shown in Table 3.

Table 3

Sample	Sample of Size $n = 5$					Sample Mean
1	36.48	39.94	42.57	39.53	33.81	38.47
2	43.13	37.97	42.41	39.61	43.30	41.28
3	41.64	39.01	37.77	38.94	41.10	39.69
4	40.37	43.49	37.60	40.14	38.88	40.10
5	38.62	33.43	45.17	42.66	39.98	39.97
6	38.98	41.35	36.80	43.56	39.92	40.12
7	42.48	37.00	35.87	39.62	38.74	38.74
8	39.38	37.02	41.60	40.34	37.62	39.19
9	42.82	45.77	35.16	42.56	39.75	41.21
10	36.19	35.20	37.74	40.46	37.47	37.41
11	36.59	41.62	42.18	39.23	39.26	39.78
12	38.57	42.13	45.39	38.22	46.18	42.10
13	38.40	39.06	43.60	31.46	37.03	37.91
14	34.29	47.73	37.27	41.82	33.33	38.89
15	42.28	43.29	37.69	37.32	40.06	40.13
16	34.31	43.58	40.02	41.13	42.99	40.41
17	38.71	39.03	39.39	42.62	38.41	39.63
18	38.63	39.66	39.47	41.13	38.01	39.38
19	39.09	33.86	37.57	41.65	35.22	37.48
20	40.94	37.50	38.72	41.64	35.48	38.86
21	38.72	35.89	37.82	35.04	37.06	36.91
22	39.64	36.30	35.54	40.40	38.74	38.12
23	38.22	38.49	33.60	40.18	39.07	37.91
24	40.93	40.53	37.55	37.30	37.16	38.69
25	33.27	38.92	37.14	39.90	33.83	36.61
26	39.44	37.28	35.70	41.97	36.80	38.24
27	38.83	41.41	38.87	39.40	37.20	39.14
28	40.10	36.96	35.73	43.00	38.11	38.78

Table 3 (cont'd)

29	41.93	36.57	37.55	35.14	38.75	37.99
30	31.25	38.85	39.25	35.07	39.77	36.84
31	38.47	34.45	30.43	41.76	41.61	37.34
32	37.98	35.56	43.97	44.96	37.81	40.06
33	43.34	40.94	35.17	41.74	37.59	39.76
34	39.80	44.44	37.53	40.52	41.95	40.85
35	41.98	42.02	40.73	40.47	36.81	40.40
36	40.98	35.08	34.61	40.78	37.26	37.74
37	35.75	40.81	40.13	35.99	36.52	37.84
38	36.39	45.97	40.59	37.64	42.42	40.60
39	36.20	35.63	37.43	38.35	34.81	36.48
40	33.58	33.87	41.60	45.10	38.68	38.57
41	31.77	38.34	41.79	37.93	40.83	38.13
42	43.03	33.12	34.98	36.58	37.78	37.10
43	35.76	35.17	42.58	39.10	41.08	38.74
44	38.44	38.45	35.93	35.32	44.60	38.55
45	44.54	41.88	35.84	42.64	42.38	41.46
46	41.89	36.81	41.83	40.24	39.28	40.01
47	38.00	40.08	35.57	34.44	39.51	37.52
48	39.92	38.05	39.96	38.04	32.11	37.62
49	36.37	38.62	32.25	41.35	40.91	37.90
50	34.38	36.65	32.97	39.93	41.34	37.05
51	40.32	39.80	41.00	38.62	38.24	39.60
52	37.95	45.26	38.67	34.96	41.13	39.59
53	36.82	42.63	41.62	39.43	37.48	39.60
54	41.63	37.65	38.58	39.03	37.53	38.88
55	37.91	37.20	38.72	36.87	45.40	39.22
56	41.05	34.01	39.11	38.23	35.74	37.63
57	42.09	45.44	35.52	39.87	37.28	40.04
58	39.31	35.79	37.82	39.15	35.57	37.53
59	41.16	39.98	41.11	39.21	39.98	40.29
60	35.68	45.60	39.34	36.65	43.30	40.11
61	36.07	39.63	42.55	41.72	36.81	39.36
62	38.97	36.83	41.01	38.12	35.27	38.04
63	33.70	39.15	34.81	34.13	39.00	36.16
64	37.19	34.69	36.21	34.34	39.07	36.30
65	33.99	44.87	42.52	40.22	39.26	40.17
66	41.40	27.62	34.57	40.08	34.65	35.66
67	40.14	34.45	38.26	38.09	39.72	38.13
68	33.64	42.62	32.08	34.30	37.34	36.00
69	35.36	39.02	43.98	41.19	32.47	38.40
70	43.26	37.85	35.82	37.11	36.22	38.05
71	36.24	38.07	33.38	38.43	39.88	37.20
72	38.55	43.06	41.07	36.58	37.02	39.26
73	41.26	36.99	36.17	38.98	36.03	37.89
74	37.31	38.41	41.18	39.76	39.64	39.26
75	36.26	41.84	42.50	37.70	41.21	39.90
76	39.27	38.61	44.53	38.08	35.01	39.10

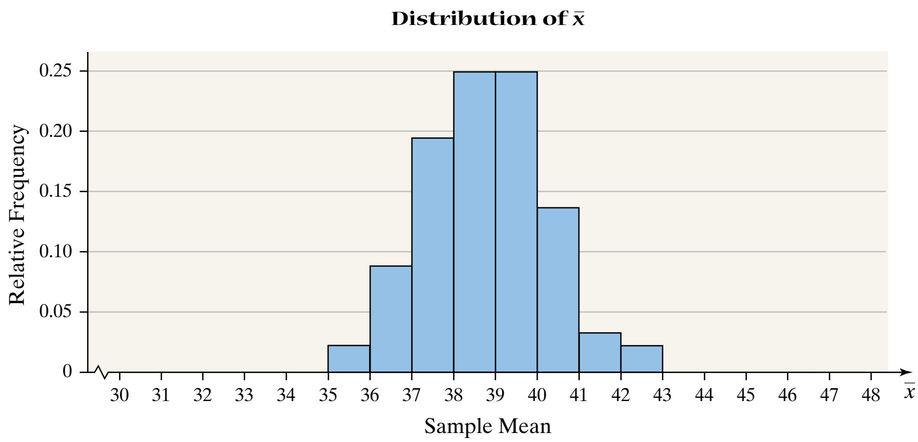
Table 3 (cont'd)

77	39.14	40.83	39.83	37.78	36.51	38.82
78	42.53	43.41	41.01	33.71	39.47	40.03
79	45.34	32.61	33.81	39.03	40.32	38.22
80	36.31	35.55	37.12	38.74	40.80	37.70
81	31.40	41.80	40.15	42.53	37.62	38.70
82	41.01	39.02	39.68	36.61	38.44	38.95
83	34.15	36.19	35.98	36.02	36.32	35.73
84	31.50	37.61	43.29	39.82	38.78	38.20
85	43.26	34.01	41.18	40.23	39.28	39.59
86	41.76	41.40	39.02	38.20	39.42	39.96
87	37.06	35.95	39.98	40.00	43.36	39.27
88	41.01	37.56	36.95	39.71	37.97	38.64
89	34.97	38.36	36.30	38.48	34.24	36.47
90	38.38	38.94	40.96	36.13	35.98	38.08
91	39.41	30.78	37.66	37.31	42.04	37.44
92	39.83	35.88	30.20	45.07	40.06	38.21
93	36.25	39.56	34.53	40.69	37.03	37.61
94	45.64	40.66	44.51	40.50	39.43	42.15
95	37.63	44.77	38.31	36.53	38.41	39.13
96	39.78	33.34	43.42	43.63	38.77	39.79
97	41.48	37.39	38.62	43.83	34.26	39.12
98	37.68	40.66	38.93	40.94	37.54	39.15
99	39.72	32.61	32.62	40.35	38.65	36.79
100	39.25	41.06	41.17	38.30	38.24	39.60

***Step 2*:** We compute the sample means for each of the 100 samples as shown in Table 3.

***Step 3*:** We draw a histogram of the 100 sample means. See Figure 2.

Figure 2



***Step 4*:** The mean of the 100 sample means is 38.72 inches, and the standard deviation is 1.374 inches.

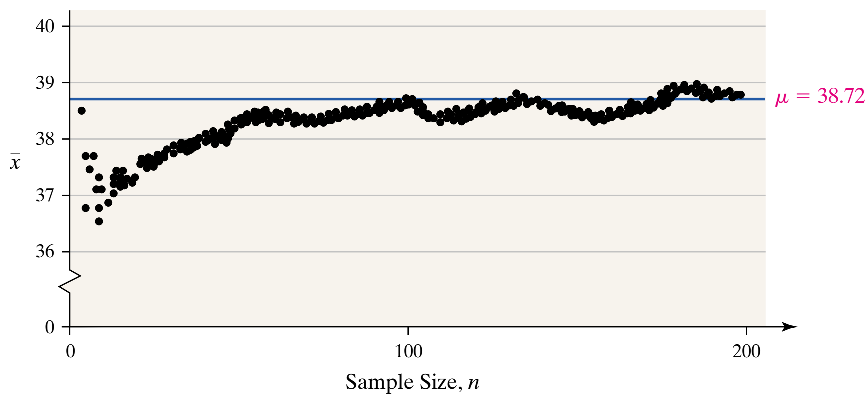
Look back at the histogram of the population data drawn in Figure 7 on page 323 from Section 7.1. Notice the center of the population distribution is the same as the center of the sampling distribution, but the spread of the population distribution is greater than that of the sampling distribution.

In Example 2 we were told that the data are approximately normal, with mean $\mu = 38.72$ inches and $\sigma = 3.17$ inches. The histogram in Figure 2 indicates that the distribution of sample means also appears to be normally distributed. In addition, the mean of the sample means is 38.72 inches, but the standard deviation is only 1.374 inches. We might conclude the following regarding the sampling distribution of $\bar{x}$.

1. *Shape*: It is normally distributed.
2. *Center*: It has mean equal to the mean of the population.
3. *Spread*: It has standard deviation less than the standard deviation of the population.

A question that we might ask is, "What role does n, the sample size, play in the sampling distribution of $\bar{x}$?" Suppose the sample mean is computed for samples of size $n = 1$ through $n = 200$. That is, the sample mean is recomputed each time an additional individual is added to the sample. The sample mean is then plotted against the sample size in Figure 3.

Figure 3



From the graph, we see that, as the sample size n increases, the sample mean gets closer to the population mean. This concept is known as the *Law of Large Numbers*.

The Law of Large Numbers

As additional observations are added to the sample, the difference between the sample mean, $\bar{x}$, and the population mean μ approaches zero.

In Other Words
As the sample size increases, the sample mean gets closer to the population mean.

So, according to the Law of Large Numbers, the more individuals we sample, the closer the sample mean gets to the population mean. This result implies that there is less variability in the distribution of the sample mean as the sample size increases. We demonstrate this result in the next example.

EXAMPLE 3 The Impact of Sample Size on Sampling Variability

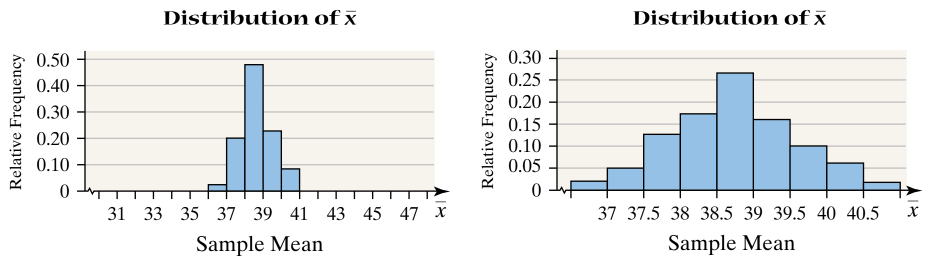
Problem: Repeat the problem in Example 2 with a sample of size $n = 15$.

Approach: The approach will be identical to that presented in Example 2, except that we let $n = 15$ instead of $n = 5$.

Solution: Figure 4(a) shows the histogram of the sample means using the same scale as Figure 2. Compare this with the histogram in Figure 2. Notice that the histogram in Figure 4(a) shows less dispersion than the histogram in Figure 2. This implies that there is less variability in the distribution of $\bar{x}$ with $n = 15$.

We redraw the histogram in Figure 4(a) using a different class width in Figure 4(b). The histogram in Figure 4(b) is symmetric and mound shaped. This is an indication that the distribution of the sample mean is approximately normally distributed. The mean of the 100 sample means is 38.72 inches (just as in Example 2); however, the standard deviation is now 0.81 inches.

Figure 4



(a) (b)

From the results of Examples 2 and 3, we conclude that, as the sample size n increases, the standard deviation of the distribution of $\bar{x}$ decreases. Although the proof is beyond the scope of this text, we should be convinced that the following result is reasonable.

The Mean and Standard Deviation of the Sampling Distribution of $\bar{x}$

Suppose that a simple random sample of size n is drawn from a large population* with mean μ and standard deviation σ. The sampling distribution of $\bar{x}$ will have mean $\mu_{\bar{x}} = \mu$ and standard deviation $\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}}$. The standard deviation of the sampling distribution of $\bar{x}$ is called the **standard error of the mean** and is denoted $\sigma_{\bar{x}}$.

In Other Words

Regardless of the distribution of the population, the sampling distribution of $\bar{x}$ will have a mean equal to the mean of the population and a standard deviation equal to the standard deviation of the population divided by the square root of the sample size!

For the population presented in Example 2, if we draw a simple random sample of size $n = 5$, the sampling distribution $\bar{x}$ will have mean $\mu_{\bar{x}} = 38.72$ inches and standard deviation

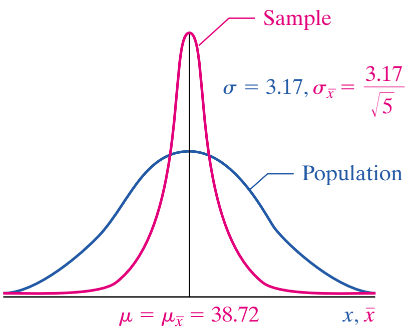
$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{3.17}{\sqrt{5}} \approx 1.418 \text{ inches}$$

Now Work Problem 11.

*Technically, we assume that we are drawing a simple random sample from an infinite population. For populations of finite size N, $\sigma_{\bar{x}} = \sqrt{\dfrac{N-n}{N-1}} \cdot \dfrac{\sigma}{\sqrt{n}}$. However, if the sample size is less than 5% of the population size ($n < 0.05N$), the effect of $\sqrt{\dfrac{N-n}{N-1}}$ (the finite population correction factor) can be ignored without affecting the results.

Now that we know how to determine the mean and standard deviation for any sampling distribution of $\bar{x}$, we can concentrate on the shape of the distribution. Refer back to Figures 2 and 4 from Examples 2 and 3. Recall that the population from which the sample was drawn was normal. The shapes of these histograms imply that the sampling distribution of $\bar{x}$ is also normal. This leads us to believe that if the population is normal the distribution of the sample mean is also normal.

Figure 5



The Shape of the Sampling Distribution of $\bar{x}$ If X Is Normal

If a random variable X is normally distributed, the distribution of the sample mean, $\bar{x}$, is normally distributed.

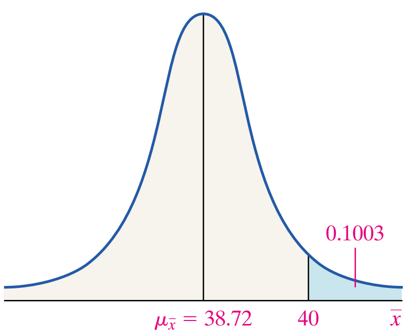
For example, the height of 3-year-old females is modeled by a normal random variable with mean $\mu = 38.72$ inches and standard deviation $\sigma = 3.17$ inches. The distribution of the sample mean, $\bar{x}$, the mean height of a simple random sample of $n = 5$ three-year-old females, is normal with mean $\mu_{\bar{x}} = 38.72$ inches and standard deviation $\sigma_{\bar{x}} = \frac{3.17}{\sqrt{5}}$ inches. See Figure 5.

EXAMPLE 4 Describing the Distribution of the Sample Mean

Problem: The height, X, of all 3-year-old females is approximately normally distributed with mean $\mu = 38.72$ inches and standard deviation $\sigma = 3.17$ inches. Compute the probability that a simple random sample of size $n = 10$ results in a sample mean greater than 40 inches. That is, compute $P(\bar{x} > 40)$.

Approach: The random variable X is normally distributed, so the sampling distribution of $\bar{x}$ will also be normally distributed. The mean of the sampling distribution is $\mu_{\bar{x}} = \mu$, and its standard deviation is $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$. We convert the random variable $\bar{x} = 40$ to a Z-score and then find the area under the standard normal curve to the right of this Z-score.

Figure 6



Solution: The sample mean is normally distributed with mean $\mu_{\bar{x}} = 38.72$ inches and standard deviation $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{3.17}{\sqrt{10}} = 1.002$ inch.

Figure 6 displays the normal curve with the area we wish to compute shaded. We convert the random variable $\bar{x} = 40$ to a Z-score and obtain

$$Z = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{\bar{x} - \mu_{\bar{x}}}{\frac{\sigma}{\sqrt{n}}} = \frac{40 - 38.72}{1.002} = 1.28$$

The area to the right of $Z = 1.28$ is $1 - 0.8997 = 0.1003$.

Interpretation: The probability of obtaining a sample mean greater than 40 inches from a population whose mean is 38.72 inches is 0.1003. That is, $P(\bar{x} \geq 40) = 0.1003$. If we take 1000 simple random samples of $n = 10$ three-year-olds from this population and if the population mean is 38.72 inches, about 100 of the samples will result in a mean height that is 40 inches or more.

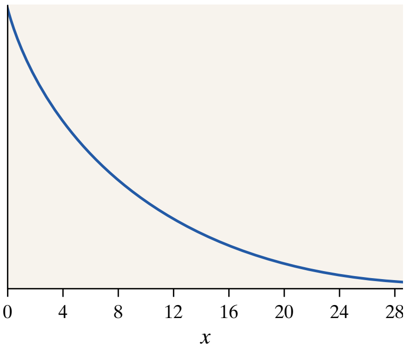
Now Work Problem 19.

3 Describe the Distribution of the Sample Mean for Samples Obtained from a Population That Is Not Normal

What if the population from which the sample is drawn is not normal?

EXAMPLE 5 Sampling from a Population That Is Not Normal

Figure 7



Problem: Figure 7 shows the graph of an *exponential density function* with mean and standard deviation equal to 10. The exponential distribution is used to model lifetimes of electronic components and to model the time required to serve a customer or repair a machine.

Clearly, the distribution of the population is not normal. Approximate the sampling distribution of $\bar{x}$ by obtaining, through simulation, 300 random samples of size (a) $n = 3$, (b) $n = 12$, and (c) $n = 30$ from the probability distribution.

Approach

Step 1: Use MINITAB, Excel, or some other statistical software to obtain 300 random samples for each sample size.

Step 2: Compute the sample mean of each of the 300 random samples.

Step 3: Draw a histogram of the 300 sample means.

Solution

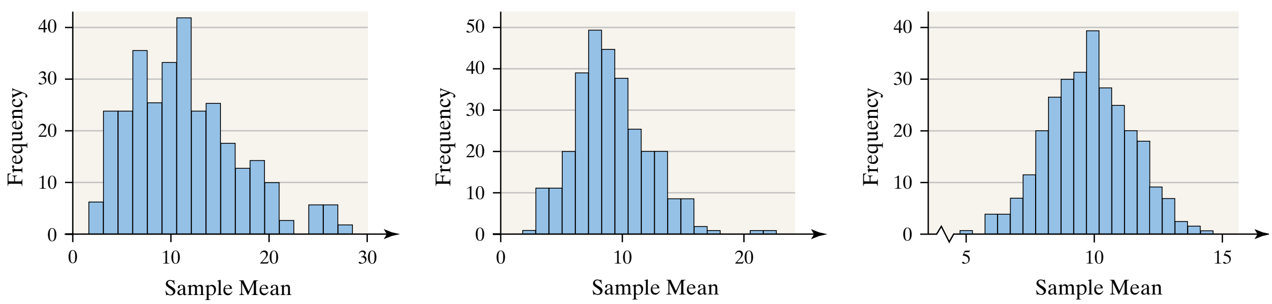
Step 1: Using MINITAB, we obtain 300 random samples of size (a) $n = 3$, (b) $n = 12$, and (c) $n = 30$. For example, in the first random sample of size $n = 30$, we obtained the following results:

9.2	20.0	17.0	2.4	2.6	19.9	21.2	5.7	8.1	10.8
1.2	22.3	18.4	4.2	9.9	41.8	4.2	1.2	10.8	2.1
11.3	17.9	28.0	12.1	3.0	0.5	4.5	14.2	5.0	11.4

Step 2: We compute the mean of each of the 300 random samples, using MINITAB. For example, the sample mean of the first sample of size $n = 30$ is 11.36.

Step 3: Figure 8(a) displays the histogram of $\bar{x}$ that results from simulating 300 random samples of size $n = 3$ from an exponential distribution with $\mu = 10$ and $\sigma = 10$. Figure 8(b) displays the histogram of $\bar{x}$ that results from simulating 300 random samples of size $n = 12$, and Figure 8(c) displays the histogram of $\bar{x}$ that results from simulating 300 random samples of size $n = 30$.

Figure 8



(a) $n = 3$ **(b)** $n = 12$ **(c)** $n = 30$

Notice that, as the sample size increases, the distribution of the sample mean becomes more normal, even though the population clearly is not normal!

We formally state the results of Example 5 as the *Central Limit Theorem*.

The Central Limit Theorem

Regardless of the shape of the population, the sampling distribution of $\bar{x}$ becomes approximately normal as the sample size n increases.

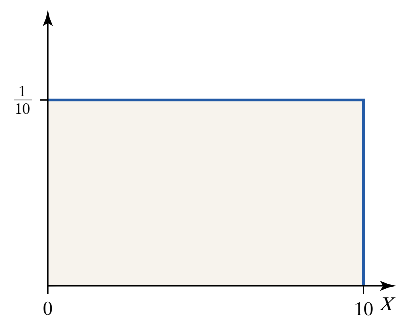
In Other Words

For any population, regardless of its shape, as the sample size increases, the shape of the distribution of the sample mean becomes more "normal."

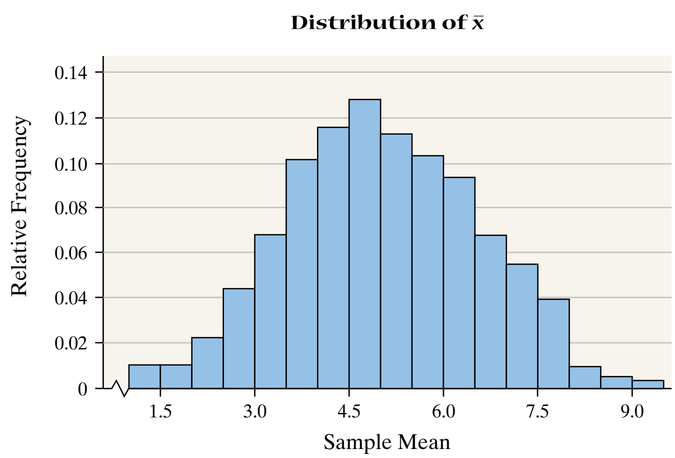
So, if the random variable X is normally distributed, the sampling distribution of $\bar{x}$ will be normal. If the sample size is large enough, the sampling distribution of $\bar{x}$ will be approximately normal, *regardless of the shape of the distribution of X*. But how large does the sample size need to be before we can say that the sampling distribution of $\bar{x}$ is approximately normal? The answer depends on the shape of the distribution of the population. Distributions that are highly skewed will require a larger sample size for the distribution of $\bar{x}$ to become approximately normal.

For example, from Example 5 we see that this right skewed distribution required a sample size of about 30 before the distribution of the sample mean is approximately normal. However, Figure 9(a) shows a uniform distribution for $0 \le X \le 10$. Figure 9(b) shows the distribution of the sample mean for $n = 3$. Figure 9(c) shows the distribution of the sample mean for $n = 12$, and Figure 9(d) shows the distribution of the sample mean for $n = 30$. Notice that even for $n = 3$ the distribution of the sample mean is approximately normal.

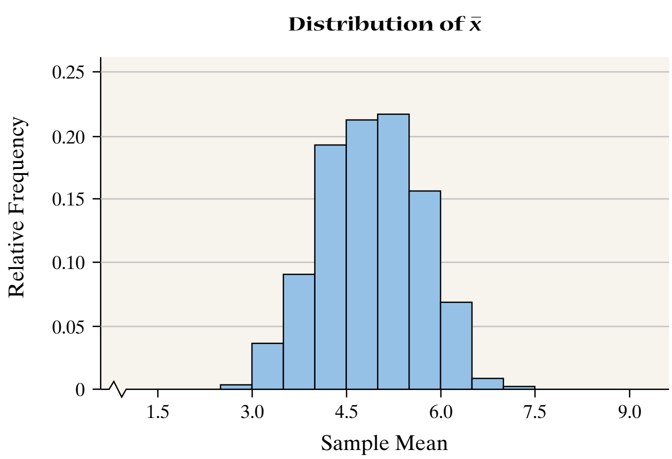
Figure 9



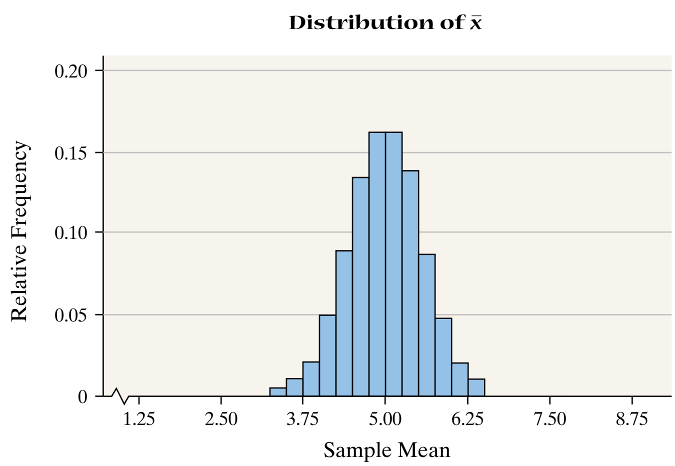
(a) Uniform Distribution



(b) Distribution of $\bar{x}$; $n = 3$



(c) Distribution of $\bar{x}$; $n = 12$



(d) Distribution of $\bar{x}$; $n = 30$

Table 4 shows the distribution of the cumulative number of children for 50- to 54-year-old mothers who had a live birth in 2002.

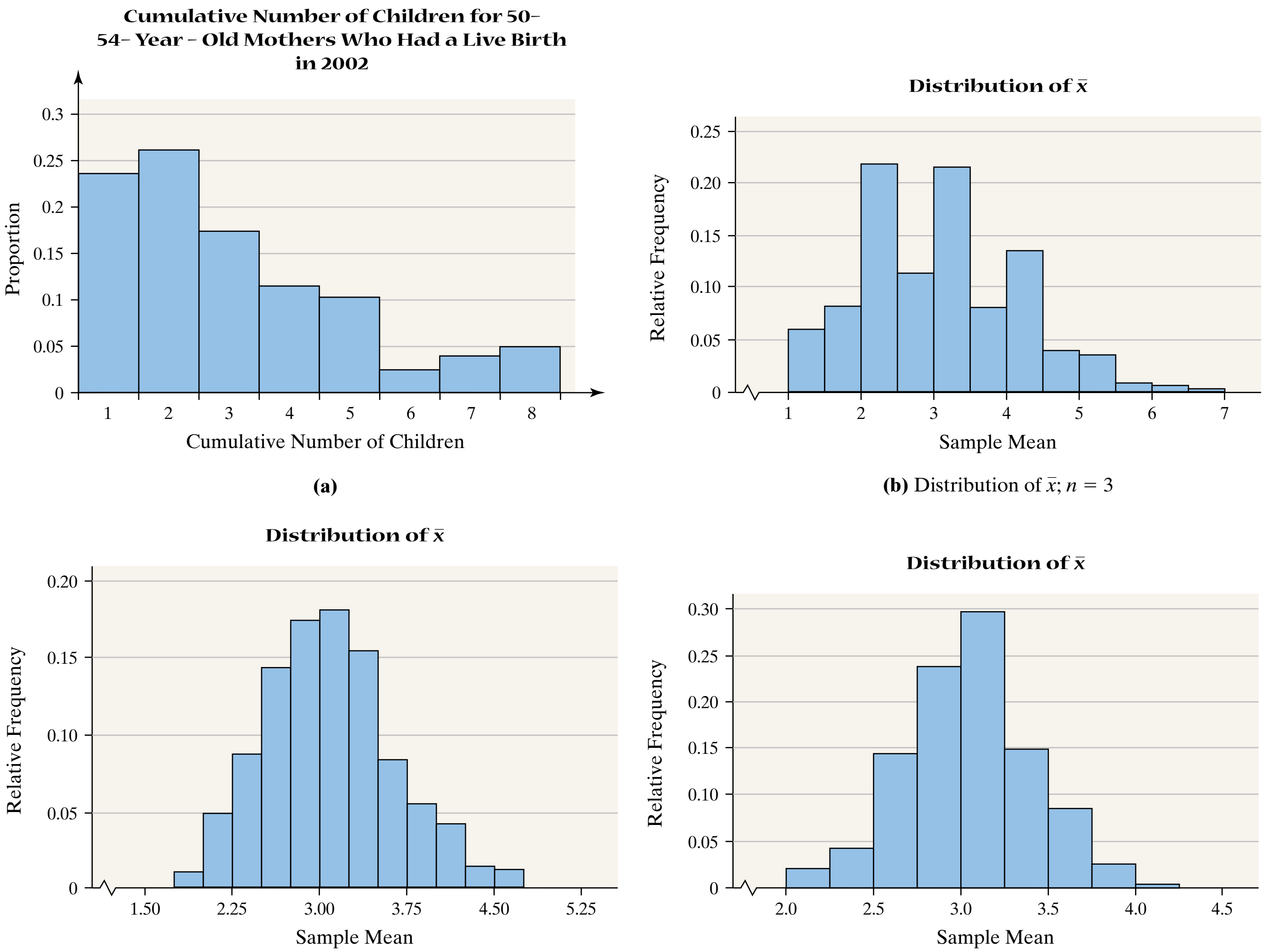
Table 4

x (number of children)	$P(x)$
1	0.241
2	0.257
3	0.172
4	0.119
5	0.103
6	0.027
7	0.031
8	0.050

Source: U.S. Census Bureau

Figure 10(a) shows the probability histogram for this distribution. Figure 10(b) shows the distribution of the sample mean number of children for a random sample of $n = 3$ mothers. Figure 10(c) shows the distribution of the sample mean number of children for a random sample of $n = 12$ mothers, and Figure 10(d) shows the distribution of the sample mean for a random sample of $n = 30$ mothers. In this instance, the distribution of the sample mean is very close to normal for $n = 12$.

Figure 10



(c) Distribution of $\bar{x}$; $n = 12$

(d) Distribution of $\bar{x}$; $n = 30$

The results of Example 5 and Figures 9 and 10 confirm that the shape of the distribution of the population dictates the size of the sample required before the distribution of the sample mean can be called normal. With that said, so that we err on the side of caution, we will say that the distribution of the sample mean is approximately normal provided that the sample size is greater than or equal to 30 if the distribution of the population is unknown or not normal.

EXAMPLE 6

Applying the Central Limit Theorem

Problem: According to the U.S. Department of Agriculture, the mean calorie intake of males 20 to 39 years old is $\mu = 2716$, with standard deviation $\sigma = 72.8$. Suppose a nutritionist analyzes a simple random sample of $n = 35$ males between the ages of 20 and 39 years old and obtains a sample mean calorie intake of $\bar{x} = 2750$ calories. What is the probability that a random sample of 35 males between the ages of 20 and 39 years old would result in a sample mean of 2750 calories or higher? Are the results of the survey unusual? Why?

Approach

Step 1: We recognize that we are computing a probability regarding a sample mean, so we need to know the sampling distribution of $\bar{x}$. Because the population from which the sample is drawn is not known to be normal, the sample size must be greater than or equal to 30 to use the results of the Central Limit Theorem.

Step 2: Determine the mean and standard deviation of the sampling distribution of $\bar{x}$.

Step 3: Convert the sample mean to a Z-score.

Step 4: Use Table IV to find the area under the normal curve.

Solution

Step 1: Because the sample size is $n = 35$, the Central Limit Theorem states that the sampling distribution of $\bar{x}$ is approximately normal.

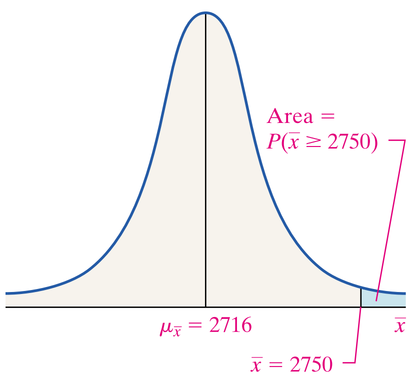
Step 2: The mean of the sampling distribution of $\bar{x}$ will equal the mean of the population, so $\mu_{\bar{x}} = 2716$. The standard deviation of the sampling distribution of $\bar{x}$ will equal the standard deviation of the population divided by the square root of the sample size, so $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{72.8}{\sqrt{35}} = 12.305$.

Step 3: We convert $\bar{x} = 2750$ to a Z-score.

$$Z = \frac{2750 - 2716}{\frac{72.8}{\sqrt{35}}} = 2.76$$

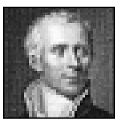
Step 4: We wish to know the probability that a random sample of $n = 35$ from a population whose mean is 2716 results in a sample mean of at least 2750. That is, we wish to know $P(\bar{x} \geq 2750)$. See Figure 11.

Figure 11



CAUTION

The Central Limit Theorem has to do only with the *shape* of the distribution of the sample mean, not with its center and spread! The mean of the distribution of $\bar{x}$ is μ and the standard deviation of $\bar{x}$ is $\frac{\sigma}{\sqrt{n}}$, regardless of the size of the sample, n.



Historical Note

Pierre Simon Laplace was born on March 23, 1749 in Normandy, France. At age 16, Laplace attended Caen University, where he studied theology. While there, his mathematical talents were discovered, which led him to Paris, where he got a job as professor of mathematics at the École Militaire. In 1773, Laplace was elected to the Académie des Sciences. Laplace was not humble. It is reported that, in 1780, he stated that he was the best mathematician in Paris. In 1799, Laplace published the first two volumes of *Méchanique céleste*, in which he discusses methods for calculating the motion of the planets. On April 9, 1810, Laplace presented the Central Limit Theorem to the Academy.

This probability is represented by the area under the standard normal curve to the right of $Z = 2.76$.

$$P(\bar{x} \geq 2750) = P(Z \geq 2.76) = 1 - 0.9971 = 0.0029$$

Interpretation: If the population mean is 2716 calories, the probability that a random sample of 35 males between the ages of 20 and 39 will result in a sample mean calorie intake of 2750 calories or higher is 0.0029. This means that fewer than 1 sample in 100 will result in a sample mean of 2750 calories or higher if the population mean is 2716 calories. We can conclude one of two things based on this result.

1. The mean number of calories for males 20 to 39 years old is 2716, and we just happened to randomly select 35 individuals who, on average, consume more calories.
2. The mean number of calories consumed by 20- to 39-year-old males is higher than 2716 calories.

In statistical inference, we are inclined to accept the second possibility as the more reasonable choice. We recognize there is a possibility that our conclusion is incorrect.

Now Work Problem 25.

Summary: Shape, Center, and Spread of the Distribution of $\bar{x}$

Shape, Center, and Spread of Population	Distribution of the Sample Mean		
	Shape	**Center**	**Spread**
Normal with mean μ and standard deviation σ	Regardless of the sample size n, the shape of the distribution of the sample mean is normal	$\mu_{\bar{x}} = \mu$	$\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}}$
Population is not normal with mean μ and standard deviation σ	As the sample size n increases, the distribution of the sample mean becomes approximately normal	$\mu_{\bar{x}} = \mu$	$\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}}$

MAKING AN INFORMED DECISION

How Much Time Do You Spend in a Day . . . ?

The American Time Use Survey is a survey of adult Americans conducted by the Bureau of Labor Statistics. The purpose of the survey is to learn how Americans allocate their time in a day. As a reporter for the school newspaper, you wish to file a report that compares the typical student at your school to the rest of America.

For those Americans who are currently attending school, the mean amount of time spent in class in a day is 5.11 hours, and the mean amount of time spent studying and doing homework is 2.50 hours. The mean amount of time Americans spend watching television each day is 2.57 hours.

Conduct a survey of 35 randomly selected full-time students at your school in which you ask the following questions:

(a) On average, how much time do you spend attending class each day?

(b) On average, how much time do you spend studying and doing homework each day?

(c) On average, how much time do you spend watching television each day? If you do not watch television, write 0 hours.

1. For each question, describe the sampling distribution of the sample mean. Use the national norms as estimates for the population means for each variable. Use the sample standard deviation as an estimate of the population standard deviation.
2. Compute probabilities regarding the values of the statistics obtained from the survey. Are any of the results unusual?

Write an article for your newspaper reporting your findings.

8.1 ASSESS YOUR UNDERSTANDING

Concepts and Vocabulary

1. Explain what a sampling distribution is.
2. State the Central Limit Theorem.
3. The standard deviation of the sampling distribution of $\bar{x}$, denoted $\sigma_{\bar{x}}$, is called the _____ _____ of the _____.
4. As the sample size increases, the difference between the sample mean, $\bar{x}$, and the population mean, μ, approaches _____.
5. What are the mean and standard deviation of the sampling distribution of $\bar{x}$, regardless of the distribution of the population from which the sample was drawn?
6. If a random sample of size $n = 6$ is taken from a population, what is required to say that the sampling distribution of $\bar{x}$ is approximately normal?
7. To cut the standard error of the mean in half, the sample size must be increased by a factor of _____.
8. *True or False*: The mean and standard deviation of the distribution of $\bar{x}$ is $\mu_{\bar{x}} = \mu$ and $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$, respectively, even if the population is not normal.
9. Suppose a simple random sample of size $n = 10$ is obtained from a population that is normally distributed with $\mu = 30$ and $\sigma = 8$. What is the sampling distribution of $\bar{x}$?
10. Suppose a simple random sample of size $n = 40$ is obtained from a population with $\mu = 50$ and $\sigma = 4$. Does the population need to be normally distributed for the sampling distribution of $\bar{x}$ to be approximately normally distributed? Why? What is the sampling distribution of $\bar{x}$?

Skill Building

In Problems 11–14, determine $\mu_{\bar{x}}$ and $\sigma_{\bar{x}}$ from the given parameters of the population and the sample size.

11. NW $\mu = 80, \sigma = 14, n = 49$
12. $\mu = 64, \sigma = 18, n = 36$
13. $\mu = 52, \sigma = 10, n = 21$
14. $\mu = 27, \sigma = 6, n = 15$
15. Suppose a simple random sample of size $n = 49$ is obtained from a population with $\mu = 80$ and $\sigma = 14$.
    (a) Describe the sampling distribution of $\bar{x}$.
    (b) What is $P(\bar{x} > 83)$?
    (c) What is $P(\bar{x} \le 75.8)$?
    (d) What is $P(78.3 < \bar{x} < 85.1)$?
16. Suppose a simple random sample of size $n = 36$ is obtained from a population with $\mu = 64$ and $\sigma = 18$.
    (a) Describe the sampling distribution of $\bar{x}$.
    (b) What is $P(\bar{x} < 62.6)$?
    (c) What is $P(\bar{x} \ge 68.7)$?
    (d) What is $P(59.8 < \bar{x} < 65.9)$?
17. Suppose a simple random sample of size $n = 12$ is obtained from a population with $\mu = 64$ and $\sigma = 17$.
    (a) What must be true regarding the distribution of the population in order to use the normal model to compute probabilities regarding the sample mean? Assuming this condition is true, describe the sampling distribution of $\bar{x}$.
    (b) Assuming the requirements described in part (a) are satisfied, determine $P(\bar{x} < 67.3)$.
    (c) Assuming the requirements described in part (a) are satisfied, determine $P(\bar{x} \ge 65.2)$.
18. Suppose a simple random sample of size $n = 20$ is obtained from a population with $\mu = 64$ and $\sigma = 17$.
    (a) What must be true regarding the distribution of the population in order to use the normal model to compute probabilities regarding the sample mean? Assuming this condition is true, describe the sampling distribution of $\bar{x}$.
    (b) Assuming the requirements described in part (a) are satisfied, determine $P(\bar{x} < 67.3)$.
    (c) Assuming the requirements described in part (a) are satisfied, determine $P(\bar{x} \ge 65.2)$.
    (d) Compare the results obtained in parts (b) and (c) with the results obtained in parts (b) and (c) in Problem 17. What effect does increasing the sample size have on the probabilities? Why do you think this is the case?

Applying the Concepts

19. **Gestation Period** NW The length of human pregnancies is approximately normally distributed with mean $\mu = 266$ days and standard deviation $\sigma = 16$ days.
    (a) What is the probability a randomly selected pregnancy lasts less than 260 days?
    (b) What is the probability that a random sample of 20 pregnancies has a mean gestation period of 260 days or less?
    (c) What is the probability that a random sample of 50 pregnancies has a mean gestation period of 260 days or less?
    (d) What might you conclude if a random sample of 50 pregnancies resulted in a mean gestation period of 260 days or less?
    (e) What is the probability a random sample of size 15 will have a mean gestation period within 10 days of the mean?

20. Serum Cholesterol As reported by the U.S. National Center for Health Statistics, the mean serum high-density-lipoprotein (HDL) cholesterol of females 20 to 29 years old is $\mu = 53$. If serum HDL cholesterol is normally distributed with $\sigma = 13.4$, answer the following questions:

(a) What is the probability that a randomly selected female 20 to 29 years old has a serum cholesterol above 60?
(b) What is the probability that a random sample of 15 female 20- to 29-year-olds has a mean serum cholesterol above 60?
(c) What is the probability that a random sample of 20 female 20- to 29-year-olds has a mean serum cholesterol above 60?
(d) What effect does increasing the sample size have on the probability? Provide an explanation for this result.
(e) What might you conclude if a random sample of 20 female 20- to 29-year-olds has a mean serum cholesterol above 60?

21. Old Faithful The most famous geyser in the world, Old Faithful in Yellowstone National Park, has a mean time between eruptions of 85 minutes. If the interval of time between eruptions is normally distributed with standard deviation 21.25 minutes, answer the following questions: (*Source*: www.unmuseum.org)

(a) What is the probability that a randomly selected time interval between eruptions is longer than 95 minutes?
(b) What is the probability that a random sample of 20 time intervals between eruptions has a mean longer than 95 minutes?
(c) What is the probability that a random sample of 30 time intervals between eruptions has a mean longer than 95 minutes?
(d) What effect does increasing the sample size have on the probability? Provide an explanation for this result.
(e) What might you conclude if a random sample of 30 time intervals between eruptions has a mean longer than 95 minutes?

22. Medical Residents In a 2003 study, the Accreditation Council for Graduate Medical Education found that medical residents work an average of 81.7 hours per week. Suppose the number of hours worked per week by medical residents is normally distributed with standard deviation 6.9 hours per week. (*Source*: www.medrecinst.com)

(a) What is the probability that a randomly selected medical resident works less than 75 hours per week?
(b) What is the probability that the mean number of hours worked per week by a random sample of five medical residents is less than 75 hours?
(c) What is the probability that the mean number of hours worked per week by a random sample of eight medical resident is less than 75 hours?
(d) What might you conclude if the mean number of hours worked per week by a random sample of eight medical residents is less than 75 hours?

23. Rates of Return in Stocks The S&P 500 is a collection of 500 stocks of publicly traded companies. Using data obtained from Yahoo!Finance, the monthly rates of return of the S&P 500 since 1950 are normally distributed. The mean rate of return is 0.007233 (0.7233%), and the standard deviation for rate of return is 0.04135 (4.135%).

(a) What is the probability that a randomly selected month has a positive rate of return? That is, what is $P(x > 0)$?
(b) Treating the next 12 months as a simple random sample, what is the probability that the mean monthly rate of return will be positive? That is, with $n = 12$, what is $P(\bar{x} > 0)$?
(c) Treating the next 24 months as a simple random sample, what is the probability that the mean monthly rate of return will be positive?
(d) Treating the next 36 months as a simple random sample, what is the probability that the mean monthly rate of return will be positive?
(e) Use the results of parts (b)–(d) to describe the likelihood of earning a positive rate of return on stocks as the investment time horizon increases.

24. Gas Mileage Based on tests of the Chevrolet Cobalt, engineers have found that the miles per gallon in highway driving are normally distributed, with a mean of 32 miles per gallon and a standard deviation 3.5 miles per gallon.

(a) What is the probability that a randomly selected Cobalt gets more than 34 miles per gallon?
(b) Suppose that 10 Cobalts are randomly selected and the miles per gallon for each car are recorded. What is the probability that the mean miles per gallon exceed 34 miles per gallon?
(c) Suppose that 20 Cobalts are randomly selected and the miles per gallon for each car are recorded. What is the probability that the mean miles per gallon exceed 34 miles per gallon? Would this result be unusual?

25. Oil Change NW The shape of the distribution of the time required to get an oil change at a 10-minute oil-change facility is unknown. However, records indicate that the mean time for an oil change is 11.4 minutes and the standard deviation for oil-change time is 3.2 minutes.

(a) To compute probabilities regarding the sample mean using the normal model, what size sample would be required?
(b) What is the probability that a random sample of $n = 40$ oil changes results in a sample mean time less than 10 minutes?

26. Time Spent in the Drive-Through The quality-control manager of a Long John Silver's restaurant wishes to analyze the length of time a car spends at the drive-through window waiting for an order. According to records obtained from the restaurants, it is determined that the mean time spent at the window is 59.3 seconds with a standard deviation of 13.1 seconds. The distribution of time at the window is skewed right (data based on information provided by Danica Williams, student at Joliet Junior College).

(a) To obtain probabilities regarding a sample mean using the normal model, what size sample is required?
(b) The quality-control manager wishes to use a new delivery system designed to get cars through the drive-through system faster. A random sample of 40 cars results in a sample mean time spent at the window of 56.8 seconds. What is the probability of obtaining a sample mean of 56.8 seconds or less assuming the population mean is 59.3 seconds? Do you think that the new system is effective?

27. Insect Fragments The Food and Drug Administration sets Food Defect Action Levels (FDALs) for some of the various foreign substances that inevitably end up in the food we eat and liquids we drink. For example, the FDAL for insect filth in peanut butter is 3 insect fragments (larvae, eggs, body parts, and so on) per 10 grams. A random sample of 50 ten-gram portions of peanut butter is obtained and results in a sample mean of $\bar{x} = 3.6$ insect fragments per ten-gram portion.

(a) Why is the sampling distribution of $\bar{x}$ approximately normal?

(b) What is the mean and standard deviation of the sampling distribution of $\bar{x}$ assuming $\mu = 3$ and $\sigma = \sqrt{3}$.

(c) Suppose a simple random sample of $n = 50$ ten-gram samples of peanut butter results in a sample mean of 3.6 insect fragments. What is the probability a simple random sample of 50 ten-gram portions results in a mean of at least 3.6 insect fragments? Is this result unusual? What might we conclude?

28. Burger King's Drive-Through Suppose cars arrive at Burger King's drive-through at the rate of 20 cars every hour between 12:00 noon and 1:00 P.M. A random sample of 40 one-hour time periods between 12:00 noon and 1:00 P.M. is selected and has 22.1 as the mean number of cars arriving.

(a) Why is the sampling distribution of $\bar{x}$ approximately normal?

(b) What is the mean and standard deviation of the sampling distribution of $\bar{x}$ assuming $\mu = 20$ and $\sigma = \sqrt{20}$.

(c) What is the probability that a simple random sample of 40 one-hour time periods results in a mean of at least 22.1 cars? Is this result unusual? What might we conclude?

29. Blows to the Head In a 2003 study of the long-term effects of concussions in football players, researchers at Virginia Tech concluded that college football players receive a mean of 50 strong blows to the head, each with an average of 40G (40 times the force of gravity). Assume the standard deviation is 16 strong blows to the head. What is the probability that a random sample of 60 college football players results in a mean of 45 or fewer strong blows to the head? Would this be unusual?
(*Source*: Neuroscience for Kids, faculty.washington.edu/chudler/nfl.html)

30. Domestic Vacation Costs According to the AAA (American Automobile Association, April 20, 2005), a family of two adults and two children on vacation in the United States will pay an average of $247.02 per day for food and lodging with a standard deviation of $60.41 per day. Suppose a random sample of 50 families of two adults and two children is selected and monitored while on vacation in the United States. What is the probability that the average daily expenses for the sample are over $260.00 per day? Would this be unusual?

31. Sampling Distributions NW The following data represent the ages of the winners of the Academy Award for Best Actor for the years 1999–2004.

2004: Jamie Foxx	37
2003: Sean Penn	43
2002: Adrien Brody	29
2001: Denzel Washington	47
2000: Russell Crowe	36
1999: Kevin Spacey	40

(a) Compute the population mean, μ.

(b) List all possible samples with size $n = 2$. There should be ${}_6C_2 = 15$ samples.

(c) Construct a sampling distribution for the mean by listing the sample means and their corresponding probabilities.

(d) Compute the mean of the sampling distribution.

(e) Compute the probability that the sample mean is within 3 years of the population mean age.

(f) Repeat parts (b)–(e) using samples of size $n = 3$. Comment on the effect of increasing the sample size.

32. Sampling Distributions The following data represent the running lengths (in minutes) of the winners of the Academy Award for Best Picture for the years 1999–2004.

2004: *Million Dollar Baby*	132
2003: *The Lord of the Rings*: *The Return of the King*	201
2002: *Chicago*	112
2001: *A Beautiful Mind*	134
2000: *Gladiator*	155
1999: *American Beauty*	120

(a) Compute the population mean, μ.

(b) List all possible samples with size $n = 2$. There should be ${}_6C_2 = 15$ samples.

(c) Construct a sampling distribution for the mean by listing the sample means and their corresponding probabilities.

(d) Compute the mean of the sampling distribution.

(e) Compute the probability that the sample mean is within 15 minutes of the population mean running times.

(f) Repeat parts (b)–(e) using samples of size $n = 3$. Comment on the effect of increasing the sample size.

33. Simulation Scores on the Stanford–Binet IQ test are normally distributed with $\mu = 100$ and $\sigma = 16$.

(a) Use MINITAB, Excel, or some other statistical software to obtain 500 random samples of size $n = 20$.

(b) Compute the sample mean for each of the 500 samples.

(c) Draw a histogram of the 500 sample means. Comment on its shape.

(d) What do you expect the mean and standard deviation of the sampling distribution of the mean to be?

(e) Compute the mean and standard deviation of the 500 sample means. Are they close to the expected values?

(f) Compute the probability that a random sample of 20 people results in a sample mean greater than 108.

(g) What proportion of the 500 random samples had a sample mean IQ greater than 108? Is this result close to the theoretical value obtained in part (f)?

34. Sampling Distribution Applet Load the sampling distribution applet on your computer. Set the applet so that the population is bell shaped. Take note of the mean and standard deviation.

APPLET

(a) Obtain 1000 random samples of size $n = 5$. Describe the distribution of the sample mean based on the results of the applet. According to statistical theory, what is the distribution of the sample mean?
(b) Obtain 1000 random samples of size $n = 10$. Describe the distribution of the sample mean based on the results of the applet. According to statistical theory, what is the distribution of the sample mean?
(c) Obtain 1000 random samples of size $n = 30$. Describe the distribution of the sample mean based on the results of the applet. According to statistical theory, what is the distribution of the sample mean?
(d) Compare the results of parts (a)–(c). How are they the same? How are they different?

35. Sampling Distribution Applet Load the sampling distribution applet on your computer. Set the applet so that the population is skewed or draw your own skewed distribution. Take note of the mean and standard deviation.

APPLET

(a) Obtain 1000 random samples of size $n = 5$. Describe the distribution of the sample mean based on the results of the applet.
(b) Obtain 1000 random samples of size $n = 10$. Describe the distribution of the sample mean based on the results of the applet.
(c) Obtain 1000 random samples of size $n = 50$. Describe the distribution of the sample mean based on the results of the applet. According to statistical theory, what is the distribution of the sample mean?
(d) Compare the results of parts (a)–(c). How are they the same? How are they different? What impact does the sample size have on the shape of the distribution of the sample mean?

8.2 Distribution of the Sample Proportion

Preparing for this Section Before getting started, review the following:

- Applications of the Normal Distribution (Section 7.3, pp. 345–349)

Objectives

1. **Describe the sampling distribution of a sample proportion**
2. **Compute probabilities of a sample proportion**

1 Describe the Sampling Distribution of a Sample Proportion

Suppose we want to determine the proportion of households in a 100-house homeowners association that favor an increase in the annual assessments to pay for neighborhood improvements. One approach that we might take is to survey all households and determine which were in favor of higher assessments. If 65 of the 100 households favor the higher assessment, the population proportion, p, of households in favor of a higher assessment is

$$p = \frac{65}{100} = 0.65$$

Of course, it is rare to gain access to all the individuals in a population. For this reason, we usually obtain estimates of population parameters such as p.

Definition

Suppose a random sample of size n is obtained from a population in which each individual either does or does not have a certain characteristic. The **sample proportion**, denoted $\hat{p}$ (read "p-hat"), is given by

$$\hat{p} = \frac{x}{n}$$

where x is the number of individuals in the sample with the specified characteristic.* The sample proportion is a statistic that estimates the population proportion, p.

*For those who studied Section 6.2 on binomial probabilities, x can be thought of as the number of successes in n trials of a binomial experiment.

EXAMPLE 1 Computing a Sample Proportion

Problem: Opinion Dynamics Corporation conducted a survey of 1000 adult Americans 18 years of age or older and asked, "Are you currently on some form of a low-carbohydrate diet?" Of the 1000 individuals surveyed, 150 indicated that they were on a low-carb diet. Find the sample proportion of individuals surveyed who are on a low-carb diet.

Approach: The sample proportion of individuals on a low-carb diet is found using the formula $\hat{p} = \frac{x}{n}$, where $x = 150$, the number of individuals in the survey with the characteristic "on a low-carb diet," and $n = 1000$.

Solution: Substituting $x = 150$ and $n = 1000$ into the formula $\hat{p} = \frac{x}{n}$, we have that $\hat{p} = \frac{150}{1000} = 0.15$. Opinion Dynamics Corporation estimates that 0.15 or 15% of adult Americans 18 years of age or older are on some form of low-carbohydrate diet.

If a second survey of 1000 American adults is conducted, it is likely the estimate of the proportion of Americans on a low-carbohydrate diet will be different because there will be different individuals in the sample. Because the value of $\hat{p}$ varies from sample to sample, it is a random variable and has a probability distribution.

To get a sense of the shape, center, and spread of the distribution of $\hat{p}$, we could repeat the exercise of obtaining simple random samples of 1000 adult Americans over and over. This would lead to a list of sample proportions. Each sample proportion would correspond to a simple random sample of 1000. A histogram of the sample proportions will give us a feel for the shape of the distribution of the sample proportion. The mean of the sample proportions will give us an idea of the center of the distribution of the sample proportion. The standard deviation of the sample proportions gives us an idea of the spread of the distribution of the sample proportions.

Rather than literally surveying 1000 adult Americans over and over again, we will use simulation to get an idea of the shape, center, and spread of the sampling distribution of the proportion.

EXAMPLE 2 Using Simulation to Describe the Distribution of the Sample Proportion

Problem: According to the Centers for Disease Control, 17% (or 0.17) of Americans have high cholestrol. Simulate obtaining 100 simple random samples of size (a) $n = 10$, (b) $n = 40$, and (c) $n = 80$. Describe the shape, center, and spread of the distribution for each sample size.

Approach: Use MINITAB, Excel, or some other statistical software package to conduct the simulation. We will perform the following steps:

Step 1: Obtain 100 simple random samples of size $n = 10$ from the population.

Step 2: Compute the sample proportion for each of the 100 samples.

Step 3: Draw a histogram of the sample proportions.

Step 4: Compute the mean and standard deviation of the sample proportions.

We then repeat these steps for samples of size $n = 40$ and $n = 80$.

Solution

Step 1: We simulate obtaining 100 simple random samples each of size $n = 10$ using MINITAB.

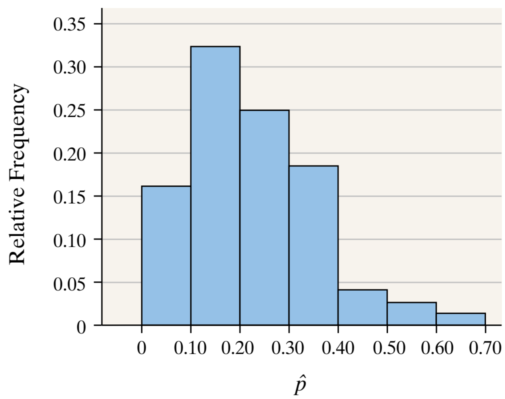
***Step 2*:** The first sample of size $n = 10$ results in none of the individuals having high cholesterol, so $\hat{p} = \frac{0}{10} = 0$. The second sample results in two of the individuals having high cholesterol, so $\hat{p} = \frac{2}{10} = 0.2$. Table 5 shows the sample proportions for all 100 simple random samples of size $n = 10$.

Table 5

0.0	0.1	0.3	0.2	0.0	0.2	0.2	0.0	0.1	0.0
0.2	0.1	0.0	0.3	0.3	0.2	0.1	0.0	0.4	0.3
0.1	0.1	0.0	0.3	0.5	0.3	0.1	0.2	0.3	0.1
0.1	0.2	0.2	0.3	0.3	0.1	0.3	0.5	0.4	0.3
0.0	0.1	0.1	0.2	0.0	0.6	0.3	0.1	0.1	0.2
0.2	0.3	0.2	0.2	0.3	0.1	0.0	0.1	0.2	0.1
0.2	0.1	0.4	0.1	0.2	0.1	0.1	0.2	0.2	0.2
0.1	0.2	0.1	0.0	0.1	0.0	0.3	0.3	0.2	0.2
0.1	0.1	0.4	0.3	0.0	0.3	0.1	0.2	0.1	0.2
0.2	0.0	0.0	0.1	0.3	0.2	0.1	0.1	0.0	0.1

***Step 3*:** Figure 12 shows a histogram of the 100 sample proportions. Notice that the shape of the distribution is skewed right.

Figure 12
Distribution of $\hat{p}$ with $n = 10$



***Step 4*:** The mean of the 100 sample proportions in Table 5 is 0.17. This is the same as the population proportion. The standard deviation of the 100 sample proportions in Table 5 is 0.1262.

We repeat Steps 1 through 4 for samples of size $n = 40$ and $n = 80$. Figure 13 shows the histogram for a sample of size $n = 40$. Notice that the shape of the distribution is skewed right (although not as skewed as the histogram with $n = 10$). The mean of the 100 sample proportions is 0.17 (the same as the population proportion), and the standard deviation is 0.0614 (less than the standard deviation for $n = 10$). Figure 14 shows the histogram for samples of size $n = 80$. Notice that the shape of the distribution is approximately normal. The mean of the 100 sample proportions for samples of size $n = 80$ is 0.17 (the same as the population proportion), and the standard deviation is 0.0408 (less than the standard deviation for $n = 40$).

We notice the following regarding the distribution of the sample proportion:

- *Shape*: As the size of the sample, n, increases, the shape of the distribution of the sample proportion becomes approximately normal.
- *Center*: The mean of the distribution of the sample proportion equals the population proportion, p.
- *Spread*: The standard deviation of the distribution of the sample proportion decreases as the sample size, n, increases.

Figure 13
Distribution of $\hat{p}$ with $n = 40$

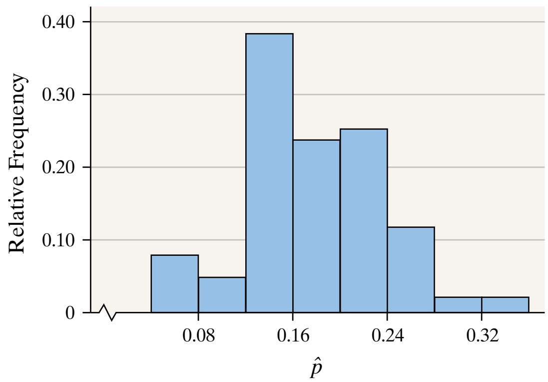
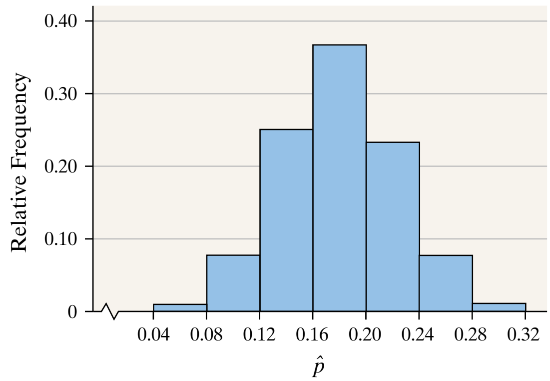


Figure 14
Distribution of $\hat{p}$ with $n = 80$



Although the proof is beyond the scope of this text, we should be convinced that the following result is reasonable.

Sampling Distribution of $\hat{p}$

For a simple random sample of size n such that $n \leq 0.05N$ (that is, the sample size is less than or equal to 5% of the population size)

- The shape of the sampling distribution of $\hat{p}$ is approximately normal provided $np(1 - p) \geq 10$.
- The mean of the sampling distribution of $\hat{p}$ is $\mu_{\hat{p}} = p$.
- The standard deviation of the sampling distribution of $\hat{p}$ is

$$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}.$$

In Other Words

The reason that the sample size cannot be more than 5% of the population size is because the success or failure of identifying an individual in the population that has the specified characteristic should not be affected by earlier observations. For example, in a population of size 100 where 14 of the individuals have brown hair, the probability that a randomly selected individual has brown hair is $14/100 = 0.14$. The probability that a second randomly selected student has brown hair is $13/99 = 0.13$. The probability changes because the sampling is done without replacement.

The condition that the sample size is no more than 5% of the population size is needed so that result obtained from one individual in the survey is independent of the result obtained from any other individual in the survey. The condition that $np(1 - p)$ is at least 10 is needed for normality.

Also, regardless of whether $np(1 - p) \geq 10$ or not, the mean of the sampling distribution of $\hat{p}$ is p and the standard deviation of the sampling distribution of $\hat{p}$ is $\sqrt{\frac{p(1-p)}{n}}$.

EXAMPLE 3 Describing the Distribution of the Sample Proportion

Problem: According to the Centers for Disease Control, 17% (or 0.17) of Americans have high cholesterol. Suppose we obtain a simple random sample of $n = 80$ Americans and determine which have high cholesterol. Describe the shape, center, and spread for the distribution of the sample proportion of Americans who have high cholesterol.

Approach: If the sample size is less than 5% of the population size and $np(1 - p)$ is at least 10, the sampling distribution of $\hat{p}$ is approximately normal, with mean $\mu_{\hat{p}} = p$ and standard deviation $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$.

Solution: There are about 295 million people in the United States. The sample of $n = 80$ is certainly less than 5% of the population size. Also, $np(1 - p) = 80(0.17)(1 - 0.17) = 11.288 \geq 10$. The distribution of $\hat{p}$ is approximately normal, with mean $\mu_{\hat{p}} = 0.17$ and standard deviation

$$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.17(1-0.17)}{80}} \approx 0.0420$$

Now Work Problem 7.

2 Compute Probabilities of a Sample Proportion

Now that we can describe the distribution of the sample proportion, we can compute probabilities of obtaining a specific sample proportion.

EXAMPLE 4 Compute Probabilities of a Sample Proportion

Problem: According to the National Center for Health Statistics, 15% of all Americans have hearing trouble.

(a) In a random sample of 120 Americans, what is the probability at least 18% have hearing trouble?

(b) Would it be unusual if a random sample of 120 Americans results in 10 having hearing trouble?

Approach: First, we determine whether the distribution of the sampling distribution is approximately normal, with mean $\mu_{\hat{p}} = p$ and standard deviation $\sigma_{\hat{p}} = \sqrt{\dfrac{p(1-p)}{n}}$, by verifying that the sample size is less than 5% of the population size and that $np(1-p) \geq 10$. Then we can use the normal distribution to determine the probabilities.

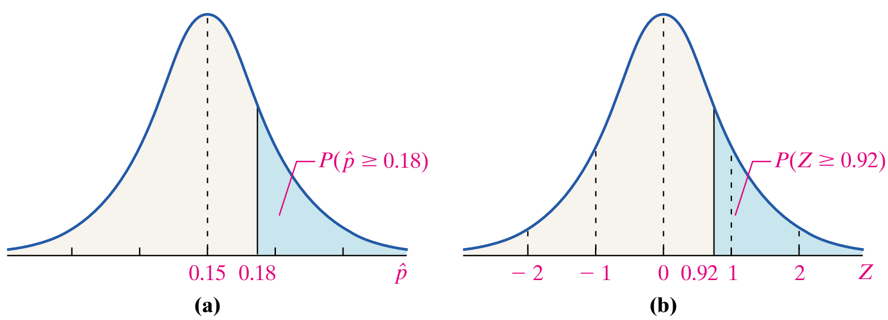
Solution: There are approximately 295 million people in the United States. The sample size of $n = 120$ is definitely less than 5% of the population size. We are told that $p = 0.15$. Because $np(1-p) = 120(0.15)(1-0.15) = 15.3 \geq 10$, the shape of the distribution of the sample proportion is approximately normal. The mean of the sample proportion $\hat{p}$ is $\mu_{\hat{p}} = 0.15$ and the standard deviation is $\sigma_{\hat{p}} = \sqrt{\dfrac{0.15(1-0.15)}{120}} \approx 0.0326$.

(a) We want to know the probability that a random sample of 120 Americans will result in a sample proportion of at least 18%, or 0.18. That is, we want to know $P(\hat{p} \geq 0.18)$. Figure 15(a) shows the normal curve with the area to the right of 0.18 shaded. To find this area, we convert $\hat{p} = 0.18$ to a standard normal random variable Z by subtracting the mean and dividing by the standard deviation. Don't forget that we round Z to two decimal places.

$$Z = \frac{\hat{p} - \mu_{\hat{p}}}{\sigma_{\hat{p}}} = \frac{0.18 - 0.15}{\sqrt{\dfrac{0.15(1-0.15)}{120}}} = 0.92$$

Figure 15(b) shows a standard normal curve with the area right of 0.92 shaded. Remember, the area to the right of $\hat{p} = 0.18$ is the same as the area to the right of $Z = 0.92$.

Figure 15



The area to the right of $Z = 0.92$ is 0.1788. Therefore,

$$P(\hat{p} \geq 0.18) = P(Z \geq 0.92) = 0.1788$$

Interpretation: The probability that a random sample of $n = 120$ Americans results in at least 18% having hearing trouble is 0.1788. This means that about 18 out of 100 random samples of size 120 will result in at least 18% having hearing trouble, even though the population proportion of Americans with hearing trouble is 0.15.

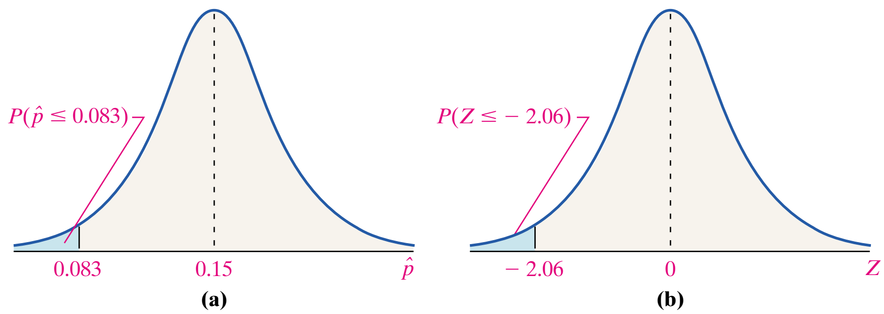
(b) A random sample of 120 Americans results in 10 having hearing trouble. The sample proportion of Americans with hearing trouble is $\hat{p} = \frac{10}{120} = 0.083$.

To determine whether a sample proportion 0.083 or less is unusual, we compute $P(\hat{p} \le 0.083)$ because if a sample proportion of 0.083 is unusual, then any sample proportion less than 0.083 is also unusual. Figure 16(a) shows the normal curve with the area to the left of 0.083 shaded. To find this area, we convert $\hat{p} = 0.083$ to a standard normal random variable Z.

$$Z = \frac{\hat{p} - \mu_{\hat{p}}}{\sigma_{\hat{p}}} = \frac{0.083 - 0.15}{\sqrt{\dfrac{0.15(1 - 0.15)}{120}}} = -2.06$$

Figure 16(b) shows a standard normal curve with the area left of -2.06 shaded. The area to the left of $\hat{p} = 0.083$ is the same as the area to the left of $Z = -2.06$.

Figure 16



The area to the left of $Z = -2.06$ is 0.0197. Therefore,

$$P(\hat{p} \le 0.083) = P(Z \le -2.06) = 0.0197$$

Interpretation: About 2 samples in 100 will result in a sample proportion of 0.083 or less from a population whose proportion is 0.15. We obtained a result that should only happen about 2 times in 100, so the results obtained are indeed unusual.

Now Work Problem 17.

8.2 ASSESS YOUR UNDERSTANDING

Concepts and Vocabulary

1. In a town of 500 households, 220 have a dog. The population proportion of dog owners in this town (expressed as a decimal) is $p =$ _____.
2. The _____ _____, denoted $\hat{p}$, is given by the formula $\hat{p} =$ _____, where x is the number of individuals with a specified characteristic in a sample of n individuals.
3. *True or False*: The population proportion and sample proportion always have the same value.
4. *True or False*: The mean of the sampling distribution of $\hat{p}$ is p.
5. Describe the circumstances under which the shape of the sampling distribution of $\hat{p}$ is approximately normal.
6. What happens to the standard deviation of $\hat{p}$ as the sample size increases? If the sample size is increased by a factor of 4, what happens to the standard deviation of $\hat{p}$?

Skill Building

In Problems 7–10, describe the sampling distribution of $\hat{p}$. Assume the size of the population is 25,000 for each problem.

7. NW $n = 500$, $p = 0.4$

8. $n = 300$, $p = 0.7$

9. $n = 1000$, $p = 0.103$

10. $n = 1010$, $p = 0.84$

11. Suppose a simple random sample of size $n = 75$ is obtained from a population whose size is $N = 10{,}000$ and whose population proportion with a specified characteristic is $p = 0.8$.
    (a) Describe the sampling distribution of $\hat{p}$.
    (b) What is the probability of obtaining $x = 63$ or more individuals with the characteristic? That is, what is $P(\hat{p} \geq 0.84)$?
    (c) What is the probability of obtaining $x = 51$ or fewer individuals with the characteristic? That is, what is $P(\hat{p} \leq 0.68)$?

12. Suppose a simple random sample of size $n = 200$ is obtained from a population whose size is $N = 25{,}000$ and whose population proportion with a specified characteristic is $p = 0.65$.
    (a) Describe the sampling distribution of $\hat{p}$.
    (b) What is the probability of obtaining $x = 136$ or more individuals with the characteristic? That is, what is $P(\hat{p} \geq 0.68)$?
    (c) What is the probability of obtaining $x = 118$ or fewer individuals with the characteristic? That is, what is $P(\hat{p} \leq 0.59)$?

13. Suppose a simple random sample of size $n = 1000$ is obtained from a population whose size is $N = 1{,}000{,}000$ and whose population proportion with a specified characteristic is $p = 0.35$.
    (a) Describe the sampling distribution of $\hat{p}$.
    (b) What is the probability of obtaining $x = 390$ or more individuals with the characteristic?
    (c) What is the probability of obtaining $x = 320$ or fewer individuals with the characteristic?

14. Suppose a simple random sample of size $n = 1460$ is obtained from a population whose size is $N = 1{,}500{,}000$ and whose population proportion with a specified characteristic is $p = 0.42$.
    (a) Describe the sampling distribution of $\hat{p}$.
    (b) What is the probability of obtaining $x = 657$ or more individuals with the characteristic?
    (c) What is the probability of obtaining $x = 584$ or fewer individuals with the characteristic?

Applying the Concepts

15. **Gardeners** Suppose a simple random sample of size $n = 100$ households is obtained from a town with 5000 households. It is known that 30% of the households plant a garden in the spring.
    (a) Describe the sampling distribution of $\hat{p}$.
    (b) What is the probability that more than 37 households in the sample plant a garden? Is this result unusual?
    (c) What is the probability that 18 or fewer households in the sample plant a garden? Is this result unusual?

16. **Text Messaging** A nationwide study in 2003 indicated that about 60% of college students with cell phones send and receive text messages with their phones. Suppose a simple random sample of $n = 1136$ college students with cell phones is obtained.
    (*Source*: promomagazine.com)
    (a) Describe the sampling distribution of $\hat{p}$, the sample proportion of college students with cell phones who send or receive text messages with their phones.
    (b) What is the probability that 665 or fewer college students in the sample send and receive text messages with their cell phones? Is this result unusual?
    (c) What is the probability that 725 or more college students in the sample send and receive text messages with their cell phone? Is this result unusual?

17. NW **Credit Cards** According to a *USA Today* "Snapshot," 26% of adults do not have any credit cards. Suppose a simple random sample of 500 adults is obtained.
    (a) Describe the sampling distribution of $\hat{p}$, the sample proportion of adults who do not have a credit card.
    (b) In a random sample of 500 adults, what is the probability that less than 24% have no credit cards?
    (c) Would it be unusual if a random sample of 500 adults results in 150 or more having no credit cards?

18. **Cell Phone Only** According to a CNN report, 7% of the population do not have traditional phones and instead rely on only cell phones. Suppose a random sample of 750 telephone users is obtained.
    (a) Describe the sampling distribution of $\hat{p}$, the sample proportion that is "cell-phone only."

(b) In a random sample of 750 telephone users, what is the probability that more than 8% are "cell-phone only"?

(c) Would it be unusual if a random sample of 750 adults results in 40 or fewer being "cell-phone only"?

19. Phishing A report released in May 2005 by First Data Corp. indicated that 43% of adults had received a "phishing" contact (a bogus e-mail that replicates an authentic site for the purpose of stealing personal information such as account numbers and passwords). Suppose a random sample of 800 adults is obtained.

(a) In a random sample of 800 adults, what is the probability that no more than 40% have received a phishing contact?

(b) Would it be unusual if a random sample of 800 adults resulted in 45% or more who had received a phishing contact?

20. Second Homes According to the National Association of Realtors, 23% of the roughly 8 million homes purchased in 2004 were considered investment properties. Suppose a random sample of 500 homes sold in 2004 is obtained.

(a) In a random sample of 500 homes sold in 2004, what is the probability that at least 125 were purchased as investment properties?

(b) Would it be unusual if a random sample of 500 homes sold in 2004 results in 20% or less being purchased as an investment property?

21. Social Security Reform A researcher studying public opinion of proposed Social Security changes obtains a simple random sample of 50 adult Americans and asks them whether or not they support the proposed changes. To say that the distribution of $\hat{p}$, the sample proportion who respond yes, is approximately normal, how many more adult Americans does the researcher need to sample if

(a) 10% of all adult Americans support the changes?

(b) 20% of all adult Americans support the changes?

22. ADHD A researcher studying ADHD among teenagers obtains a simple random sample of 100 teenagers aged 13 to 17 and asks them whether or not they have ever been prescribed medication for ADHD. To say that the distribution of $\hat{p}$, the sample proportion who respond no, is approximately normal, how many more teenagers aged 13 to 17 does the researcher need to sample if

(a) 90% of all teenagers aged 13 to 17 have never been prescribed medication for ADHD?

(b) 95% of all teenagers aged 13 to 17 have never been prescribed medication for ADHD?

23. Simulation The following exercise is meant to illustrate the normality of the distribution of the sample proportion, $\hat{p}$.

(a) Using MINITAB or some other statistical spreadsheet, randomly generate 2000 samples of size 765 from a population with $p = 0.3$. Store the number of successes in a column called x.

(b) Determine $\hat{p}$ for each of the 2000 samples by computing $\frac{x}{765}$. Store each $\hat{p}$ in a column called *phat*.

(c) Draw a histogram of the 2000 estimates of p. Comment on the shape of the distribution.

(d) Compute the mean and standard deviation of the sampling distribution of $\hat{p}$ in the simulation.

(e) Compute the theoretical mean and standard deviation of the sampling distribution of $\hat{p}$. Compare the theoretical results to the results of the simulation. Are they close?

24. The Sampling Distribution Applet Load the sampling distribution applet on your computer. Set the applet so that the population is binary with probability of success equal to 0.2.

APPLET

(a) Obtain 1000 random samples of size $n = 5$. Describe the distribution of the sample proportion based on the results of the applet.

(b) Obtain 1000 random samples of size $n = 30$. Describe the distribution of the sample proportion based on the results of the applet.

(c) Obtain 1000 random samples of size $n = 100$. Describe the distribution of the sample proportion based on the results of the applet.

(d) Compare the results of parts (a)–(c). How are they the same? How are they different?

25. Finite Population Correction Factor In this section, we assumed that the sample size was less than 5% of the size of the population. When sampling without replacement from a finite population in which $n > 0.05N$, the standard deviation of the distribution of $\hat{p}$ is given by

$$\sigma_{\hat{p}} = \sqrt{\frac{\hat{p}(1 - \hat{p})}{n - 1} \cdot \left(\frac{N - n}{N}\right)}$$

where N is the size of the population. Suppose a survey is conducted at a college having an enrollment of 6,502 students. The student council wants to estimate the percentage of students in favor of establishing a student union. In a random sample of 500 students, it was determined that 410 were in favor of establishing a student union.

(a) Obtain the sample proportion, $\hat{p}$, of students surveyed who favor establishing a student union.

(b) Calculate the standard deviation of the sampling distribution of $\hat{p}$.

Consumer Reports® Tanning Salons

Medical groups have long warned about the dangers of indoor tanning. The American Medical Association and the American Academy of Dermatology unsuccessfully petitioned the Food and Drug Administration in 1994 to ban cosmetic-tanning equipment. Three years later, the Federal Trade Commission warned the public to beware of advertised claims that "unlike the sun, indoor tanning will not cause skin cancer or skin aging" or that you can "tan indoors with absolutely no harmful side effects."

In February 1999, still under pressure from the medical community, the FDA announced that current recommendations "may allow higher exposures" to UV radiation "than are necessary." The agency proposed reducing recommended exposures and requiring simpler wording on consumer warnings. But it has not yet implemented either of these changes. An FDA spokeswoman told us that "the agency decided to postpone amendment of its standard pending the results of ongoing research and discussions with other organizations."

To make matters worse, only about half the states have any rules for tanning parlors. In some of these states, the regulation is minimal and may not require licensing, inspections, training, record keeping, or parental consent for minors. Despite this, nearly 30 million Americans, including a growing number of teenage girls, are expected to visit a tanning salon in 2005.

In a recent survey of 296 indoor-tanning facilities around the country, to our knowledge the first nationwide survey of its kind, we found evidence of widespread failures to inform customers about the possible risks, including premature wrinkling and skin cancer, and to follow recommended safety procedures, such as wearing eye goggles. Many facilities made questionable claims about indoor tanning: that it's safer than sunlight, for example, and is well controlled.

(a) In designing this survey, why is it important to sample a large number of facilities? And why is it important to sample these facilities in multiple cities?

(b) Given the fact that there are over 150,000 tanning facilities in the United States, is the condition for independence of survey results satisfied? Why?

(c) Sixty-seven of the 296 tanning facilities surveyed stated that "tanning in a salon is the same as tanning in the sun with respect to causing skin cancer." Assuming that the true proportion is 25%, describe the sampling distribution of $\hat{p}$, the sample proportion of tanning facilities that state "tanning in a salon is the same as tanning in the sun with respect to causing skin cancer." Calculate the probability that less than 22.6% of randomly selected tanning salon facilities would state that tanning in a salon is the same as tanning in the sun with respect to causing skin cancer.

(d) Forty-two of the 296 tanning facilities surveyed stated "tanning in a salon does not cause wrinkled skin." Assuming that the true proportion is 18%, describe the sampling distribution of $\hat{p}$, the sample proportion of tanning facilities that state that "tanning in a salon does not cause wrinkled skin." Calculate the probability that at least 14.2% will state that tanning in a salon does not cause wrinkled skin. Would it be unusual for 50 or fewer facilities to state that tanning in a salon does not cause wrinkled skin?

***Note to Readers**: In many cases, our test protocol and analytical methods are more complicated than described in this example. The data and discussion have been modified to make the material more appropriate for the audience.*

CHAPTER 8 Review

Summary

This chapter forms the bridge between probability and statistical inference. In Section 8.1, we discussed the distribution of the sample mean. We learned that the mean of the distribution of the sample mean equals the mean of the population ($\mu_{\bar{x}} = \mu$) and that the standard deviation of the distribution of the sample mean is the standard deviation of the population divided by the square root of the sample size $\left(\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}\right)$. If the sample is obtained from a population that is known to be normally distributed, the shape of the distribution of the sample mean is also normal. If the sample is obtained from a population that is not normal, the shape of the distribution of the sample mean becomes approximately normal as the sample size increases. This result is known as the Central Limit Theorem.

In Section 8.2, we discussed the distribution of the sample proportion. We learned that the mean of the distribution of the sample proportion is the population proportion ($\mu_{\hat{p}} = p$) and that the standard deviation of the distribution of the sample proportion is $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$. If $np(1-p) \geq 10$, then the shape of the distribution of $\hat{p}$ is approximately normal.

Formulas

Mean and Standard Deviation of the Sampling Distribution of $\bar{x}$

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

Sample Proportion

$$\hat{p} = \frac{x}{n}$$

Mean and Standard Deviation of the Sampling Distribution of $\hat{p}$

$$\mu_{\hat{p}} = p \quad \text{and} \quad \sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$$

Standardizing a Normal Random Variable

$$Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \quad \text{or} \quad Z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$$

Vocabulary

Statistical inference (p. 375)
Sampling distribution of the sample mean (p. 375)
Law of Large Numbers (p. 381)
Standard error of the mean (p. 382)
Central Limit Theorem (p. 385)
Sample proportion (p. 392)
Sampling distribution of $\hat{p}$ (p. 395)

Objectives

Section	You should be able to . . .	Example	Review Exercises
8.1	1 Understand the concept of a sampling distribution (p. 375)	1	1
	2 Describe the distribution of the sample mean for samples obtained from normal populations (p. 377)	2 through 4	2, 4, 5, 6
	3 Describe the distribution of the sample mean for samples obtained from a population that is not normal (p. 384)	5 and 6	7, 8, 13, 14
8.2	1 Describe the sampling distribution of a sample proportion (p. 392)	2 and 3	3, 4, 9(a)–12(a)
	2 Compute probabilities of a sample proportion (p. 396)	4	9(b), (c); 10(b), (c); 11(b), (c); 12(b), (c)

Review Exercises

1. In your own words, explain what a sampling distribution is.
2. Under what conditions is the sampling distribution of $\bar{x}$ normal?
3. Under what conditions is the sampling distribution of $\hat{p}$ approximately normal?
4. What is the mean and standard deviation of the sampling distribution of $\bar{x}$? What is the mean and standard deviation of the sampling distribution of $\hat{p}$?
5. **Energy Needs during Pregnancy** Suppose the total energy need during pregnancy is normally distributed with mean $\mu = 2600$ kcal/day and standard deviation $\sigma = 50$ kcal/day.
   (*Source*: American Dietetic Association)
   (a) What is the probability that a randomly selected pregnant woman has an energy need of more than 2625 kcal/day? Is this result unusual?
   (b) Describe the sampling distribution of $\bar{x}$, the sample mean daily energy requirement for a random sample of 20 pregnant women.
   (c) What is the probability that a random sample of 20 pregnant women has a mean energy need of more than 2625 kcal/day? Is this result unusual?
6. **Battery Life** The charge life of a certain lithium ion battery for camcorders is normally distributed with mean 90 minutes and standard deviation 35 minutes.
   (a) What is the probability that a battery of this type, randomly selected, lasts more than 100 minutes on a single charge? Is this result unusual?
   (b) Describe the sampling distribution of $\bar{x}$, the sample mean charge life for a random sample of 10 such batteries.
   (c) What is the probability that a random sample of 10 such batteries has a mean charge life of more than 100 minutes? Is this result unusual?
7. **Copper Tubing** A machine at K&A Tube & Manufacturing Company produces a certain copper tubing component in a refrigeration unit. The tubing components produced by the manufacturer have a mean diameter of 0.75 inch with standard deviation 0.004 inch. The quality-control inspector takes a random sample of 30 components once each week and calculates the mean diameter of these components. If the mean is either less than 0.748 inch or greater than 0.752 inch, the inspector concludes that the machine needs an adjustment.
   (a) Describe the sampling distribution of $\bar{x}$, the sample mean diameter, for a random sample of 30 such components.
   (b) What is the probability that, based on a random sample of 30 such components, the inspector will conclude that the machine needs an adjustment?
8. **Filling Machines** A machine used for filling plastic bottles with a soft drink has a known standard deviation of $\sigma = 0.05$ liter. The target mean fill volume is $\mu = 2.0$ liters.
   (a) Describe the sampling distribution of $\bar{x}$, the sample mean fill volume, for a random sample of 45 such bottles.

(b) What is the probability that a random sample of 45 such bottles has a mean fill volume that is less than 1.995 liters?
(c) What is the probability that a random sample of 45 such bottles has a mean fill volume that is more than 2.015 liters? Is this result unusual? What might we conclude?

9. Entrepreneurship A Gallup survey in March 2005 indicated that 72% of 18- to 29-year-olds, if given a choice, would prefer to start their own business rather than work for someone else. Suppose a random sample of 600 18- to 29-year-olds is obtained.
(a) Describe the sampling distribution of $\hat{p}$, the sample proportion of 18- to 29-year-olds who would prefer to start their own business.
(b) In a random sample of 600 18- to 29-year-olds, what is the probability that no more than 70% would prefer to start their own business?
(c) Would it be unusual if a random sample of 600 18- to 29-year-olds resulted in 450 or more who would prefer to start their own business?

10. Smokers According to the National Center for Health Statistics (2004), 22.4% of adults are smokers. Suppose a random sample of 300 adults is obtained.
(a) Describe the sampling distribution of $\hat{p}$, the sample proportion of adults who smoke.
(b) In a random sample of 300 adults, what is the probability that at least 50 are smokers?
(c) Would it be unusual if a random sample of 300 adults results in 18% or less being smokers?

11. Advanced Degrees According to the U.S. Census Bureau, roughly 9% of adults aged 25 years or older have an advanced degree. Suppose a random sample of 200 adults aged 25 years or older is obtained.
(a) Describe the sampling distribution of $\hat{p}$, the sample proportion of adults aged 25 years or older who have an advanced degree.
(b) In a random sample of 200 adults aged 25 years or older, what is the probability that no more than 6% have an advanced degree?
(c) Would it be unusual if a random sample of 200 adults aged 25 years or older results in 25 or more having an advanced degree?

12. Peanut and Tree Nut Allergies Peanut and tree nut allergies are considered to be the most serious food allergies. According to the National Institute of Allergy and Infectious Diseases, roughly 1% of Americans are allergic to peanuts or tree nuts. Suppose a random sample of 1500 Americans is obtained.
(a) Describe the sampling distribution of $\hat{p}$, the sample proportion of Americans allergic to peanuts or tree nuts.
(b) In a random sample of 1500 Americans, what is the probability that more than 1.5% are allergic to peanuts or tree nuts?
(c) Would it be unusual if a random sample of 1500 Americans results in fewer than 10 with peanut or tree nut allergies?

13. Principals' Salaries According to the *National Survey of Salaries and Wages in Public Schools*, the mean salary paid to public high school principals in 2004–2005 was $71,401. Assume the standard deviation is $26,145. What is the probability that a random sample of 100 public high school principals has an average salary under $65,000?

14. Teaching Supplies According to the National Education Association, public school teachers spend an average of $443 of their own money each year to meet the needs of their students. Assume the standard deviation is $175. What is the probability that a random sample of 50 public school teachers spends an average of more than $400 each year to meet the needs of their students?

THE CHAPTER 8 CASE STUDY IS LOCATED ON THE CD THAT ACCOMPANIES THIS TEXT.