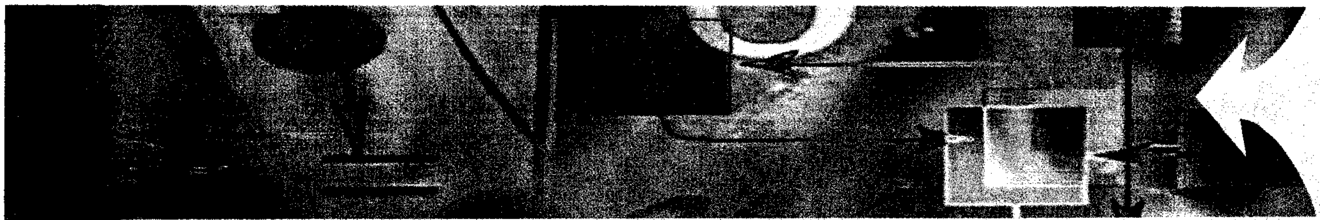


11

Inferential Statistics

P&G to Cut Spending on Local Television Advertising

Procter & Gamble Co. was dissatisfied with the "poor quality" of Nielsen Media Research's TV diaries. Because of this dissatisfaction, the company slashed its spending for local TV, or "spot," advertising. For example, in October 1997, the company spent $16 million on spot TV, but in October 1998, the company spent only $10 million, a 37.5% decrease. Nielsen Media Research used diaries to measure TV ratings in most of the nation's local television markets. But in April 1998, a high-ranking Procter & Gamble executive warned the annual convention of the Television Bureau of Advertising that "people meters" were needed in local markets, or spot television advertising would decrease. Those warnings were not heeded.

Thomas F. Craig/Index Stock Imagery

The preceding story is factual. Estimating the size of a television audience is important business. Television stations and networks can get companies to pay good money for advertising time only because people are watching television when the ads are shown. As indicated above, in local television markets, the audience size is estimated on the basis of diaries kept by a sample of households in the local viewing area. For nationwide television ratings, Nielsen Media Research previously used diaries, but switched to using people meters in 1986. The people meters are now connected to televisions in 5000 households. As they watch TV, the viewers in those households press buttons on their people meter to indicate their presence. The people meter records the gender and age of each viewer, as well as the time spent watching each channel. The people meters are also connected to the telephone line, and the data gathered are transferred automatically to Nielsen's computer every night. The fact that only 5000 well-chosen households can accurately reflect the television viewing habits of a nation of about 288 million seems remarkable.

WILLIAM SEALY GOSSETT (1876–1937) studied chemistry and mathematics at Oxford University. In 1899, after graduating at age 23, Gossett was hired by Guinness as a chemist. The Guinness Brewing Company in Dublin is one of the oldest breweries in the world. Guinness began as a family business in 1759, and its markets have grown worldwide since then. The Guinness corporation was interested in making their brewing process scientific. This was a novel idea that required new techniques, and it was why Gossett was hired. One question concerned finding the best kind of barley to use. Gossett gathered agricultural data and other information about barley. He realized that differences found in the data could be accidental or simply due to natural variation. On the other hand, they could be the result of differences in treatment or process and thus lead to better methods of brewing. There was no way to tell which was which, so new statistical methods were needed to answer that question. Gossett developed many of those methods, some while working at Guinness and some during a year spent at University College, London. Because of his employment at Guinness, he published his ideas under the pseudonym "Student." Gossett's work created the modern field of statistical inference, and the primary method for working with small samples is known as Student's t-test.

Source: gap.dcs.stand.ac.uk/%7EHistory/PictDisplay/Gosset.html

RONALD A. FISHER (1890–1962) became a statistician in 1919 at the Rothamsted Experimental Station in Great Britain. He conducted field studies and worked on genetics; in the process, he pioneered the use of randomization in experimental design and invented formal statistical methods for the analysis of experimental data. Fisher's research produced 55 groundbreaking papers that extended and clarified the revolutionary work of Gossett, Pearson, and others. This research formed the basis for the theory of making inferences from samples, which we use today.

Source: groups.dcs.stand.ac.uk/~History/PictDisplay/Gosset.html

One of the most important applications of statistical methods in the 20th century was to answer the question "Does smoking cause lung cancer?" One approach would be an experimental study involving nonsmokers who would be separated into two groups, one of whom would be required to smoke. For ethical reasons, such studies were not done. By the late 1950s, all other types of studies involving smokers and nonsmokers did, however, come to the same conclusion: a significantly higher rate of lung cancer existed among smokers.

Fisher helped advance the development of statistical methods, yet he argued against one of the most important findings made using these methods. Namely, Fisher repeatedly pointed out that the evidence against tobacco as a health hazard was only circumstantial.

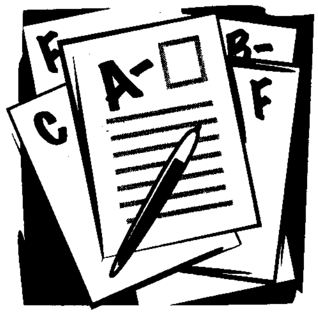
OBJECTIVES

In this chapter, you will learn:

1. why only 5000 households are sufficient to guarantee accurate Nielsen ratings;
2. how to calculate measures of central tendency and dispersion that help describe sets of data; and
3. what "margin of error" means and why it usually accompanies reports such as those on polling data.

11.1 Normal Distributions

INITIAL PROBLEM



After returning exams to her large class of 90 students, Professor LaStat reports that the mean score on the test was 74 and the standard deviation was 8. At the students' request, she agrees to "curve" the test scores. She says that curving the scores will mean that all students whose scores are at least 1.5 standard deviations above the class mean will receive A's. Similarly, all students whose scores are at least 1.5 standard deviations below the class mean will receive F's. If Professor LaStat curves the grades in this manner, about how many students in the class will get A's? About how many will get F's?

A solution of this Initial Problem is on page 711.

In Chapter 9, we discussed populations and various methods of taking a sample from a population. One of the most important applications of statistics is making predictions about an entire population based on information from a sample. This process is known as **statistical inference**. For example, public opinion polls using a sample of only about one thousand voters can accurately predict the outcome of a presidential election in which almost a hundred million votes are cast. In this chapter, you will see how such statistical inferences can be made. Because we are interested in making predictions about large populations, we will first develop a mathematical model that is often applied to large sets of data, such as the data that would be obtained from the U.S. census.

DISTRIBUTIONS OF LARGE DATA SETS

In Section 8.1, we explored ways to represent data visually. We saw that a histogram can graphically and concisely represent a large quantity of data. For large data sets, the histogram often takes on a characteristic shape as the heights of the bars increase to a maximum height and then decrease. In those instances, we can sketch a smooth curve that roughly outlines the histogram. The histogram itself can be approximated by the area of the region under that curve, as illustrated in Figure 11.1.

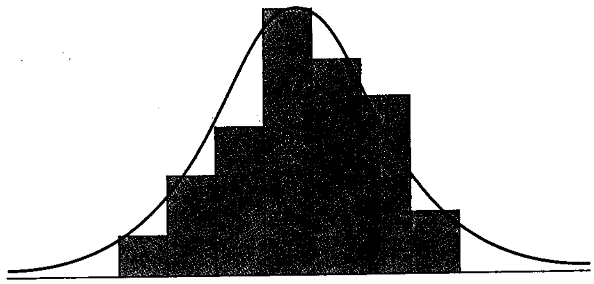


Figure 11.1

Notice that the total area of all of the bars is approximately the same as the total area under the smooth curve.

The larger the data set and the smaller the intervals, the better the smooth curve will approximate the histogram (Figure 11.2).

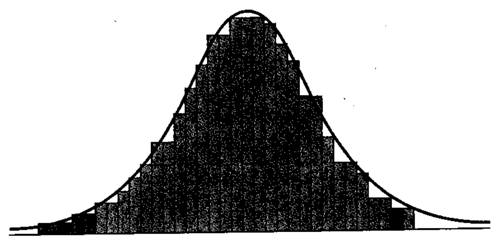


Figure 11.2

Notice again that the total area of all of the bars is about the same as the total area under the curve. However, in this case, the difference between the two areas is smaller; there is less "white space" between the curve and the histogram.

When the region under a smooth curve is used to model the histogram of a large data set, the area of the entire region under the curve represents 100% of the data. The fraction of the area under a particular part of the curve equals the fraction of the data in that interval. For example, suppose that the region under the curve in Figure 11.3 represents the histogram of a large data set.

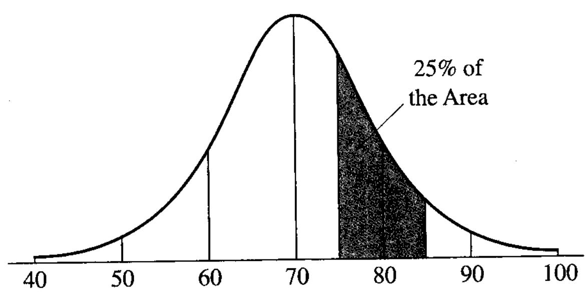


Figure 11.3

If we know that 25% of the total area under the curve lies between 75 and 85, then we know that 25% of the data itself is between 75 and 85.

Because the total area under the curve represents 100% of the data and because 100% = 1, we think of the area under the curve as one unit of area. In the case of Figure 11.3, the area of the shaded portion in the interval from 75 to 85 is 0.25 = 25% of the total area. The fact that the area under the curve is exactly 1, or 100%, will allow us to determine what percentage of data lies within a certain interval, as the next example illustrates.

EXAMPLE 11.1 Suppose the distribution of weights (in pounds) of a large number of college-age men is represented by the curve in Figure 11.4. Areas under the curve in certain intervals are indicated.

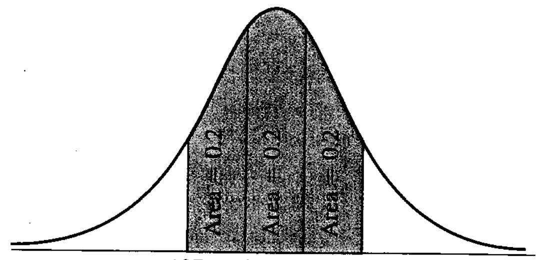


Figure 11.4

What percentage of these college-age men have weights that are

a. greater than 167 pounds and less than 192 pounds?

b. greater than 137 pounds and less than 192 pounds?

c. greater than 137 pounds and less than 222 pounds?

SOLUTION

a. From Figure 11.4, the area under the curve from 167 to 192 is 0.2. Thus, 0.2 = 20% of the men weigh more than 167 pounds and less than 192 pounds.

b. The area under the curve in the interval from 137 to 167 is 0.2. When that 0.2 is added to the 0.2 that represents the area from 167 to 192, the result is 0.4, which is the area under the curve between 137 and 192. Thus, 0.4, or 40%, of the men weigh more than 137 pounds and less than 192 pounds.

c. The region under the curve from 137 to 222 is composed of three regions, each with area 0.2. Thus, $0.2 + 0.2 + 0.2 = 0.6$, or 60%, of the men weigh more than 137 pounds and less than 222 pounds. ■

Tidbit

Only 5 months after the signing of the Declaration of Independence, on December 5, 1776, the Phi Beta Kappa (ΦBK) honor society was founded by five students at the College of William and Mary in Williamsburg, Virginia. Uppercase Greek letters have been used to name college fraternities and sororities ever since. Greek letters are also often used in mathematics to represent unknown quantities. A few commonly used lowercase Greek letters are the following:

μ (mu) represents the population mean;
σ (sigma) represents the population standard deviation;
π (pi) represents the ratio of the circumference of a circle to the diameter;
θ (theta) denotes angles.

NORMAL DISTRIBUTIONS

The shape of the curve that represents a large set of data depends on the characteristics of the data being studied. Often data can be modeled with a symmetric, bell-shaped curve like the one shown in Figure 11.4. Data that can be represented by this type of ideal, bell-shaped curve are said to have a **normal distribution** or to be **normally distributed.** The population mean and the population standard deviation, which we discussed in Chapter 9, determine the exact shape and position of the bell-shaped curve representing a data set. The "peak" of the bell-shaped curve is always directly above the population mean, μ (mu). Varying the value of μ shifts the curve to the left or to the right. The four bell-shaped curves in Figure 11.5 illustrate this relationship between μ and the position of the peak of the curve.

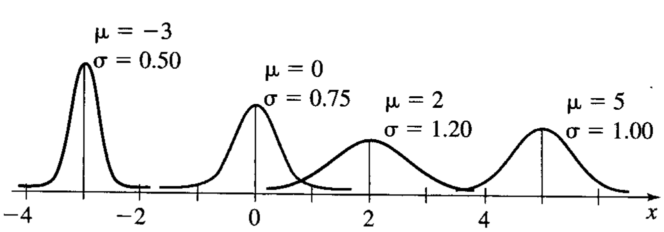


Figure 11.5

Notice that the distribution with the largest mean ($\mu = 5$) is farthest to the right, while the distribution with the smallest mean ($\mu = -3$) is farthest to the left.

Although the location of the peak of a bell-shaped curve is determined by the value of μ, neither the actual height of the peak nor the width of the curve depend on μ. Instead, the population standard deviation, σ (sigma), determines these characteristics of the graph; as we saw in Chapter 9, σ describes the "spread" of the data. If we assume that the horizontal and vertical scales remain unchanged, then small values of σ result in a curve that is tall and thin, while large values of σ make the curve short and wide.

Notice in Figure 11.5 that the distribution with the smallest standard deviation ($\sigma = 0.50$) corresponds to the narrowest graph, the one with the highest peak. The distribution with the largest standard deviation ($\sigma = 1.20$) corresponds to the widest graph, the one with the shortest peak. In the Extended Problems for this section, we will explore the effect of increasing or decreasing the standard deviation σ on the peak height and on the width of a bell-shaped curve.

EXAMPLE 11.2

a. Which bell-shaped curve in Figure 11.6 represents the data set with the smallest mean? With the largest mean?

b. Which bell-shaped curve in Figure 11.6 represents the data set with the smallest standard deviation? With the largest standard deviation?

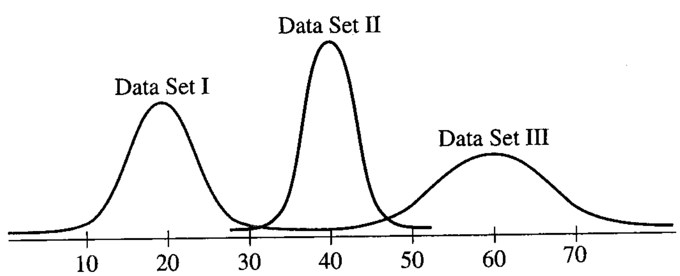


Figure 11.6

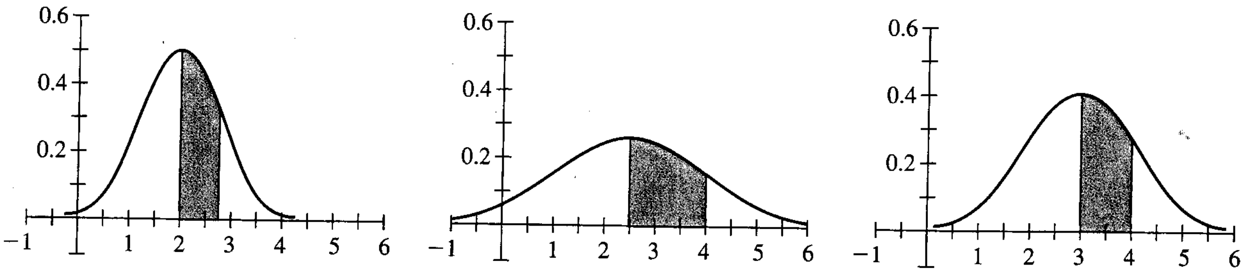
SOLUTION

a. Remember that the peak of the bell-shaped curve is located at the mean. Data set I has the smallest mean because the peak of the curve is farthest to the left. Data set III has the largest mean because the peak of the curve is the farthest to the right.

b. Remember that the standard deviation determines the height of the peak and the spread, or width, of the curve. Data set II has the smallest standard deviation because the curve is the tallest and thinnest. Data set III has the largest standard deviation because the curve is the shortest and widest. ■

THE STANDARD NORMAL DISTRIBUTION

The graphs of all normal distributions have similar bell shapes. Furthermore, the areas under the curves of all normal distributions are related in a way that will prove very useful. Figure 11.7 shows three normal distributions, including the mean and the standard deviation for each. In each normal distribution, the shaded area highlights the region from the mean to the number that is 1 standard deviation larger than the mean.



(a) Normal Distribution with $\mu = 2$ and $\sigma = 0.8$ Shaded area ≈ 0.34
$\mu + \sigma = 2 + 0.8 = 2.8$

(b) Normal Distribution with $\mu = 2.5$ and $\sigma = 1.5$ Shaded area ≈ 0.34
$\mu + \sigma = 2.5 + 1.5 = 4$

(c) Normal Distribution with $\mu = 3$ and $\sigma = 1$ Shaded area ≈ 0.34
$\mu + \sigma = 3 + 1 = 4$

Figure 11.7

Although the shapes of the three graphs differ, the shaded area under each curve is approximately 0.34; that is, 34% of the data in each distribution lies in the interval between the mean and the mean plus 1 standard deviation. It turns out that for *any* normal distribution, approximately 34% of the data will always lie in this interval. This feature of normal distributions gives us a way to determine what percentage of data lies within a certain interval. For each endpoint of the interval, we need only consider how many standard deviations above or below the mean that number is.

We have seen that the area between the mean and 1 standard deviation above the mean in any normal distribution is always approximately 0.34. Likewise, the area between 1 standard deviation above the mean and 2 standard deviations above the mean will be the same for any normal distribution. Similarly, other areas related to the mean and standard deviation are the same for any normal distribution. This fact tells us that we can pick one particular normal distribution to be the standard representative for all normal distributions. We call this special normal distribution the standard normal distribution, and it has the characteristics described next.

> ***Definition***
>
> **THE STANDARD NORMAL DISTRIBUTION**
>
> The normal distribution with a mean 0 ($\mu = 0$) and standard deviation 1 ($\sigma = 1$) is called the **standard normal distribution.**

Figure 11.8 shows the standard normal distribution.

The Standard Normal Distribution

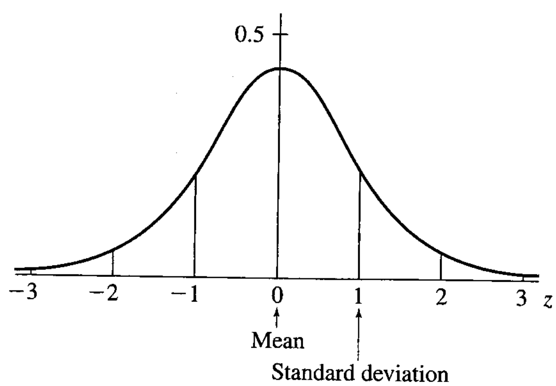


Figure 11.8

The height at the peak is approximately 0.4. The horizontal axis is marked off in units of 1 standard deviation each. It is common to use different scales on the horizontal and vertical axes to improve the appearance of the graph. The letter z is generally used instead of x for the horizontal axis to distinguish the standard normal distribution from other normal distributions.

The reason for designating a *standard* normal distribution is that all normal distributions have similar characteristics. The areas under the standard normal distribution equal the corresponding areas under any other nonstandard normal distribution. Figure 11.9 illustrates the relationship between the standard normal distribution and a nonstandard normal distribution with mean μ and standard deviation σ.

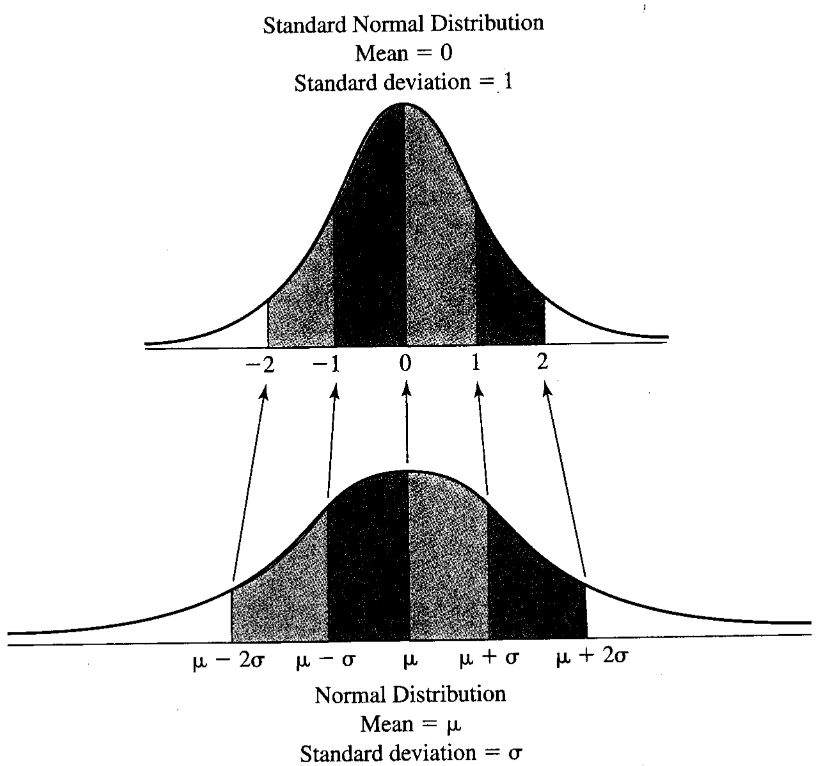


Figure 11.9

Note, for example, that the point 2 on the horizontal axis of the standard normal distribution corresponds to the point $\mu + 2\sigma$ on the horizontal axis of the nonstandard normal distribution. Similarly shaded areas in the two graphs are equal, although their shapes differ.

AREAS UNDER THE STANDARD NORMAL DISTRIBUTION CURVE

Now that we know how areas under normal distribution curves are related, we can focus our attention on finding areas under the standard normal distribution curve. As we saw earlier, when a data set is represented by a normal distribution, the fraction of data that lies between a and b is equal to the area under the normal distribution curve in the interval from a to b. The same is true for the standard normal distribution.

One way to find the area of a particular region under the standard normal distribution curve is to use a table of values, as the next two examples illustrate.

EXAMPLE 11.3 What fraction of the total area under the standard normal distribution curve lies between $a = -0.5$ and $b = 1.5$? This region is shaded in Figure 11.10.

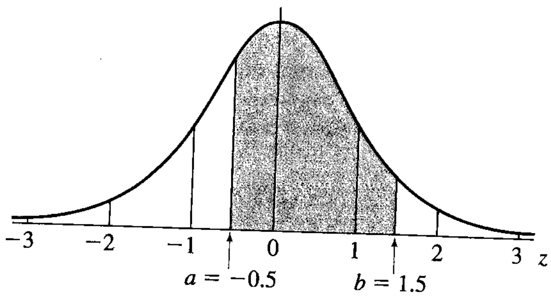


Figure 11.10

SOLUTION In Figure 11.10, the shaded region is a fraction of the area under the entire curve. This fraction of the area represents the fraction of the data in the standard normal distribution between the data points a and b. To find what fraction of the total area corresponds to the shaded region is, we will use the table of values in Table 11.1.

Table 11.1

AREA UNDER THE STANDARD NORMAL DISTRIBUTION CURVE BETWEEN a AND b

a \ b	3.0	2.5	2.0	1.5	1.0	0.5	0.0	−0.5	−1.0	−1.5	−2.0	−2.5	−3.0
−3.0	.9974	.9925	.9760	.9319	.8400	.6902	.4987	.3072	.1574	.0655	.0214	.0049	.0000
−2.5	.9925	.9876	.9711	.9270	.8351	.6853	.4938	.3023	.1525	.0606	.0165	.0000	
−2.0	.9760	.9711	.9546	.9105	.8186	.6688	.4773	.2858	.1360	.0441	.0000		
−1.5	.9319	.9270	.9105	.8664	.7745	.6247	.4332	.2417	.0919	.0000			
−1.0	.8400	.8351	.8186	.7745	.6826	.5328	.3413	.1498	.0000				
−0.5	.6902	.6853	.6688	.6247	.5328	.3830	.1915	.0000					
0.0	.4987	.4938	.4773	.4332	.3413	.1915	.0000						
0.5	.3072	.3023	.2858	.2417	.1498	.0000							
1.0	.1574	.1525	.1360	.0919	.0000								
1.5	.0655	.0606	.0441	.0000									
2.0	.0214	.0165	.0000										
2.5	.0049	.0000											
3.0	.0000												

To use Table 11.1 to find the area under the standard normal curve between a and b, locate the row in Table 11.1 corresponding to a and the column in the table corresponding to b. The fraction of the total area that lies between a and b is the number shown in the cell in the row and column you have located. In this example, $a = -0.5$ and $b = 1.5$. The cell in the row corresponding to -0.5 and in the column corresponding to 1.5 contains the number 0.6247. This cell has been shaded in Table 11.1. The value shown tells us that 62.47% of the area under the curve lies between -0.5 and 1.5. This seems reasonable, given the size of the shaded area in Figure 11.10. This area can be interpreted in other ways, too. It means that 62.47% of the data in the standard normal distribution is between $a = -0.5$ and $b = 1.5$. It also means that if a data point was selected at random, the probability is 62.47% that it would fall in the interval from -0.5 to 1.5. Finally, it means that 62.47% of the data in *any* normal distribution will lie between the point that is 0.5 standard deviation below the mean and the point that is 1.5 standard deviations above the mean. ■

EXAMPLE 11.4 What percentage of the data in a standard normal distribution lies between 0.5 and 2.5?

SOLUTION The cell in the row corresponding to $a = 0.5$ and in the column corresponding to $b = 2.5$ contains the number 0.3023. Thus, 30.23% of the data in a standard normal distribution lies in the interval from 0.5 to 2.5. ■

SYMMETRY AND TABLES FOR THE NORMAL DISTRIBUTION

Because the standard normal distribution curve is symmetric about the vertical line through 0, much of the information listed in Table 11.1 is redundant. The symmetry of the curve about $z = 0$ tells us that the fraction of the data in a standard normal distribution that lies between a and b equals the fraction of the data that lies between $-b$ and $-a$. For instance, the fraction of the data between $a = 0.0$ and $b = 1.0$ equals the fraction of the data between -1.0 and 0.0, as shown by the shaded regions in Figure 11.11.

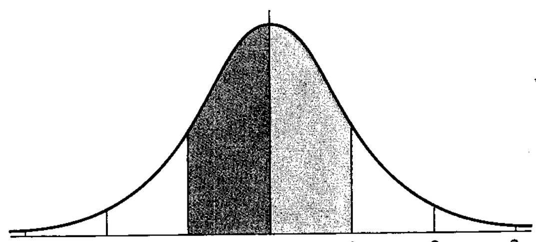


Figure 11.11

A portion of Table 11.1 is shown next (Table 11.2).

Table 11.2

EQUAL AREAS UNDER THE STANDARD NORMAL DISTRIBUTION CURVE

		b						
		3.0	2.5	2.0	1.5	1.0	0.5	0.0
a	−3.0	.9974	.9925	.9760	.9319	.8400	.6902	.4987
	−2.5	.9925	.9876	.9711	.9270	.8351	.6853	.4938
	−2.0	.9760	.9711	.9546	.9105	.8186	.6688	.4773
	−1.5	.9319	.9270	.9105	.8664	.7745	.6247	.4332
	−1.0	.8400	.8351	.8186	.7745	.6826	.5328	.3413
	−0.5	.6902	.6853	.6688	.6247	.5328	.3830	.1915
	0.0	.4987	.4938	.4773	.4332	.3413	.1915	.0000

Tidbit

Tables have been used to organize numerical information for thousands of years. Cuneiform tablets from as early as 1600 B.C. contain tables for multiplication, reciprocals, and square roots. Babylonians were mathematically advanced and applied mathematics to the study of astronomy. The use of tables made astronomical calculations easier. As recently as the 1960s, before calculators were widely available, tables were regularly used to find square roots and logarithms. Today, tables are used for frequently needed numerical data, such as areas under the standard normal curve.

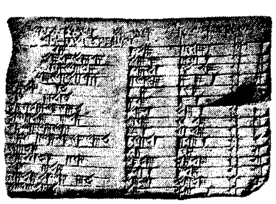


Table of Pythagorean Triples Cuneiform Tablet, 1700 B.C.

The equality of the shaded regions in Figure 11.11 is also illustrated by the equality of the entries at the ends of the arrow in Table 11.2. Notice the other pairs of equal entries in the table.

Most tables for the standard normal distribution contain no redundant data like the repeated values in Table 11.2. Instead, they require the user of the table to add and subtract entries and to use the symmetry of the curve to calculate areas not listed in the table. Table 11.3 shows such a table of areas for the standard normal distribution curve.

Table 11.3

STANDARD NORMAL DISTRIBUTION AREAS

z	Area Above Interval 0 to z	z	Area Above Interval 0 to z	z	Area Above Interval 0 to z
0.1	0.0398	1.1	0.3643	2.1	0.4821
0.2	0.0793	1.2	0.3849	2.2	0.4861
0.3	0.1179	1.3	0.4032	2.3	0.4893
0.4	0.1554	1.4	0.4192	2.4	0.4918
0.5	0.1915	1.5	0.4332	2.5	0.4938
0.6	0.2257	1.6	0.4452	2.6	0.4953
0.7	0.2580	1.7	0.4554	2.7	0.4965
0.8	0.2881	1.8	0.4641	2.8	.04974
0.9	0.3159	1.9	0.4713	2.9	0.4981
1.0	0.3413	2.0	0.4772	3.0	0.4987

The values in the table give the area under a standard normal distribution curve for the interval from 0 to z. A reference book containing more extensive tables for the standard normal distribution is *Standard Mathematical Tables*, published by the CRC Press.

To illustrate how to use Table 11.3, suppose we know that measurements from a population have a standard normal distribution. By finding the entry for $z = 0.5$ in Table 11.3, we know that the area under the standard normal distribution curve from 0 to 0.5 is 0.1915. We conclude that 19.15% of the measurements of our population have values between 0 and 0.5, as shown in Figure 11.12.

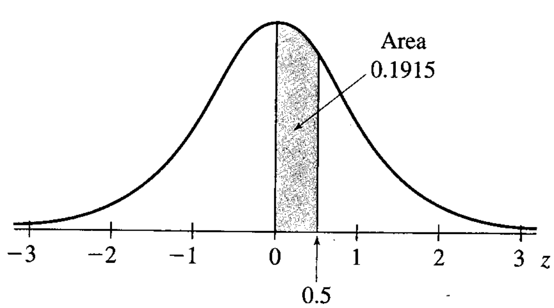


Figure 11.12

The next example shows how to combine areas from Table 11.3 to find a desired area.

EXAMPLE 11.5 Suppose the measurements from a population have a standard normal distribution. Find the percentage of the measurements that lie between $z = -1.8$ and $z = 1.3$.

SOLUTION From Table 11.3 we see that the area under the standard normal distribution curve from $z = 0$ to $z = 1.3$ is 0.4032. Similarly, we find that the area under the standard normal distribution curve from $z = 0$ to $z = 1.8$ is 0.4641. By symmetry, the area from $z = -1.8$ to $z = 0$ is the same as the area from 0 to 1.8. Thus, the area under the curve from $z = -1.8$ to $z = 1.3$ is given by $0.4641 + 0.4032 = 0.8673$. This area is illustrated in Figure 11.13.

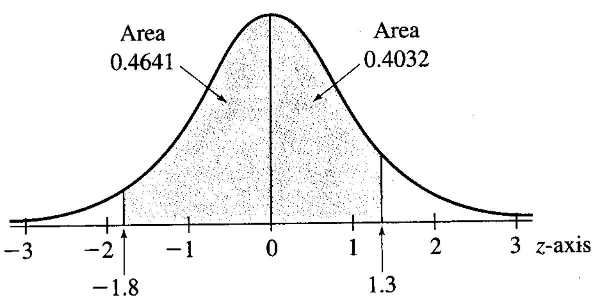


Figure 11.13 ■

We conclude that 86.73% of the measurements lie between $z = -1.8$ and $z = 1.3$. In Example 11.5, we added areas to find the desired area under the standard normal distribution curve. Sometimes we must subtract one area from another to find the desired area, as the next example illustrates.

EXAMPLE 11.6 Suppose the measurements on a population have a standard normal distribution. Find the percentage of the measurements that lie between $z = 1.2$ and $z = 1.7$.

SOLUTION From Table 11.3 we see that the area under the standard normal distribution from $z = 0$ to $z = 1.7$ is 0.4554. Similarly, we find that the area from $z = 0$ to $z = 1.2$ is 0.3849. The area that lies above the interval from $z = 1.2$ to $z = 1.7$ must be the area from $z = 0$ to $z = 1.7$ *minus* the area from $z = 0$ to $z = 1.2$; that is, the area is $0.4554 - 0.3849 = 0.0705$. Thus, we see that 7.05% of the measurements lie between $z = 1.2$ and $z = 1.7$. The corresponding region is shaded in Figure 11.14.

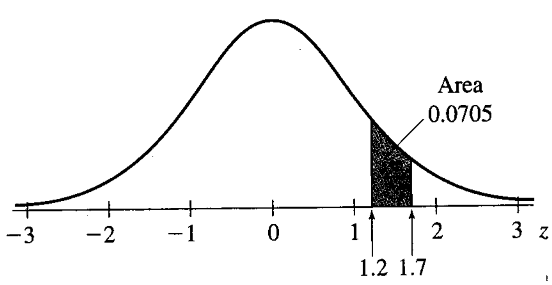
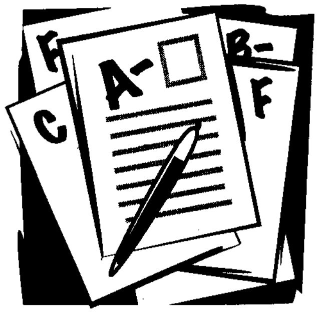


Figure 11.14 ■

SOLUTION OF THE INITIAL PROBLEM



After returning exams to her large class of 90 students, Professor LaStat reports that the mean score on the test was 74 and the standard deviation was 8. At the students' request, she agrees to "curve" the test scores. She says that curving the scores will mean that all students whose scores are at least 1.5 standard deviations above the class mean will receive A's. Similarly, all students whose scores are at least 1.5 standard deviations below the class mean will receive F's. If Professor LaStat curves the grades in this manner, about how many students in the class will get A's? About how many will get F's?

SOLUTION Because the class is large, it is likely that the test scores have a normal distribution. (Large data sets are often normally distributed.) The percentage of scores in any given range of scores may be determined by comparing this normal distribution to the standard normal distribution. If the scores are "curved" as described, then the mean score of 74 will correspond to a score of 0 in the standard normal distribution. A score that is 1.5 deviations above the mean is $74 + 1.5 \times 8 = 86$ and corresponds to a score of 1.5 in the standard normal distribution.

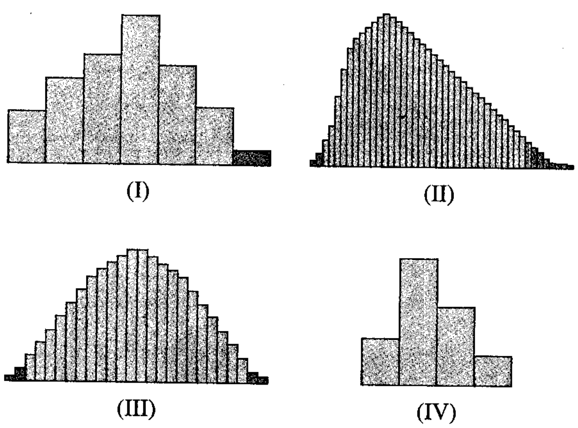
To determine the percentage of students who will receive A's, we can look at the standard normal distribution rather than the actual normal distribution of test scores. To find the percentage of A's, we must determine the area under the standard normal curve to the right of $z = 1.5$. From Table 11.3, we see that approximately 43.32% of the scores will fall between 0 and 1.5. By the symmetry of the bell-shaped curve, 50% of the scores in the standard normal distribution are greater than 0 and 50% are less than 0. Thus, approximately $50\% - 43.32\% = 6.68\%$ of the scores will be greater than 1.5, which means that approximately $0.0668 \times 90 \approx 6$ students will get A's.

By symmetry, the number of students who will receive F's will also be approximately $0.0668 \times 90 \approx 6$.

PROBLEM SET 11.1

Problems 1 and 2

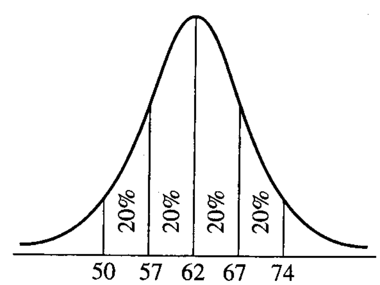
Refer to the following histograms.



1. Which of histograms (I)–(IV) could best be approximated by the region under a smooth curve? Explain.
2. Which of histograms (I)–(IV) could best be approximated by the region under a smooth normal distribution curve? Explain.

Problems 3 through 6

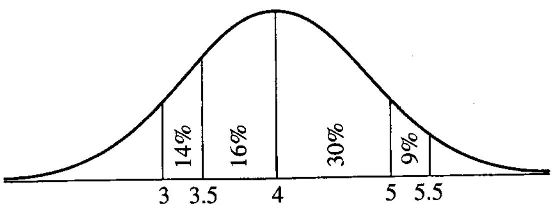
Suppose the following figure represents the distribution of the weights (in pounds) of a certain large breed of dog.



3. What percentage of the dogs in this population weigh between 67 and 74 pounds?
4. What percentage of the dogs in this population weigh between 57 and 62 pounds?
5. What percentage of the dogs in this population weigh between 50 and 67 pounds?
6. What percentage of the dogs in this population weigh between 50 and 74 pounds?

Problems 7 through 10

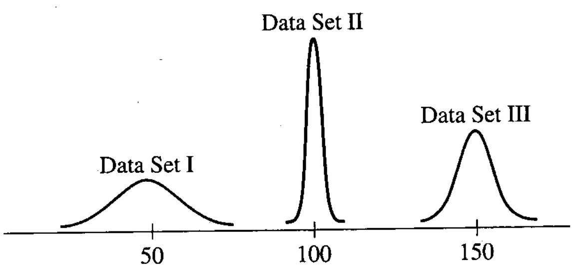
Suppose the following figure represents the distribution of body lengths (in inches) from a large population of a certain species of hamster.



7. What percentage of the hamsters in this population are between 3.5 and 4 inches long?
8. What percentage of the hamsters in this population are between 5 and 5.5 inches long?
9. What percentage of the hamsters in this population are between 3 and 5.5 inches long?
10. What percentage of the hamsters in this population are between 3.5 and 5 inches long?

Problems 11 and 12

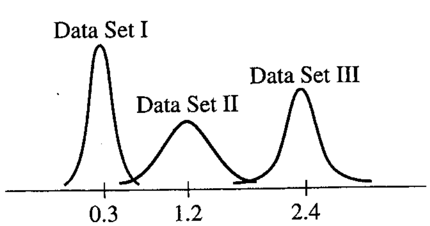
Refer to the following figure, where three normal distributions are sketched on the same x-axis.



11. Which data set has the largest mean? Which has the smallest mean? How can you tell?
12. Which data set has the largest standard deviation? Which has the smallest standard deviation? How can you tell?

Problems 13 and 14

Refer to the following figure, where three normal distributions are sketched on the same x-axis.

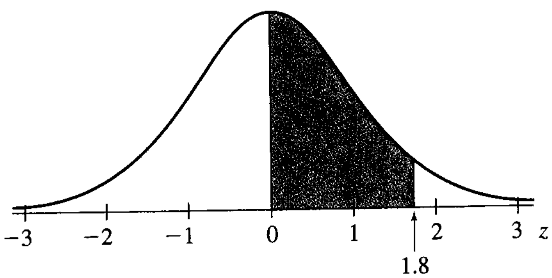


13. Which data set has the largest standard deviation? Which has the smallest standard deviation? How can you tell?
14. Which data set has the largest mean? Which has the smallest mean? How can you tell?
15. Sketch examples of two normal distributions that have the same mean, but have different standard deviations. Clearly label your graphs.
16. Sketch examples of two normal distributions that have different means, but have the same standard deviations. Clearly label your graphs.

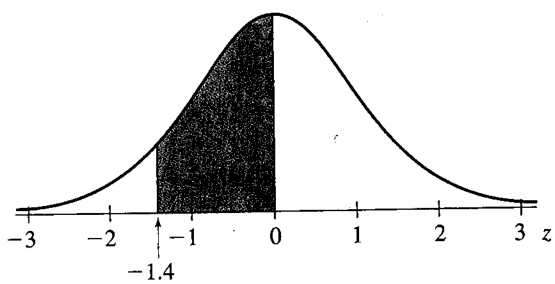
Problems 17 through 22

For the following standard normal curve, find the area of the shaded region using Table 11.3 and interpret the results.

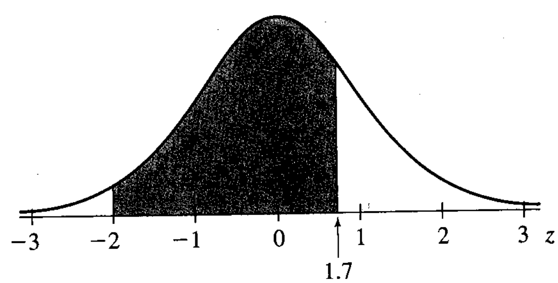
17.



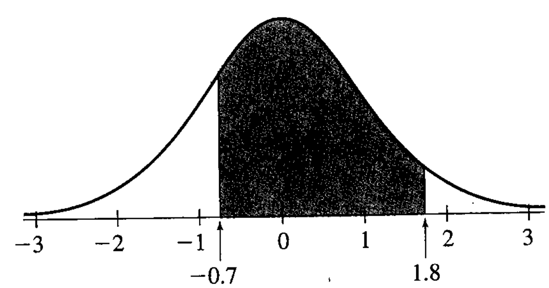
18.



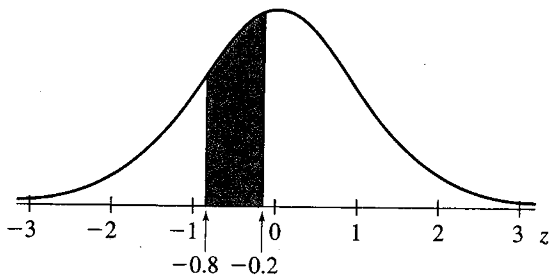
19.



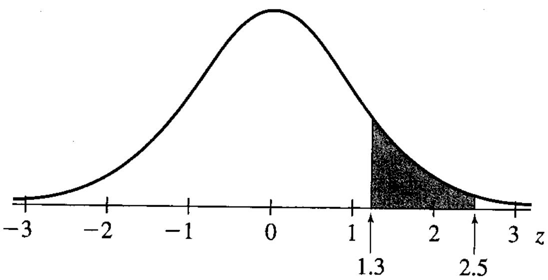
20.



21.



22.



23. Use Table 11.3 to find the area under the standard normal curve between 0 and 1.5. Then sketch the standard normal curve and shade the appropriate region.

24. Use Table 11.3 to find the area under the standard normal curve between −0.7 and 0. Then sketch the standard normal curve and shade the appropriate region.

25. Use Table 11.3 to find the area under the standard normal curve between −2.1 and 1.9. Then sketch the standard normal curve and shade the appropriate region.

26. Use Table 11.3 to find the area under the standard normal curve between 0.6 and 1.8. Then sketch the standard normal curve and shade the appropriate region.

27. Consider measurements taken from a population that has a standard normal distribution.
   - **a.** Find the percentage of the data that have a value between 1 and 3.
   - **b.** Find the percentage of the data that have a value larger than 2.
   - **c.** Find the percentage of the data that are not between −1 and 1.

28. Consider measurements taken from a population that has a standard normal distribution.
   - **a.** Find the percentage of the data that have a value between −2 and 3.
   - **b.** Find the percentage of the data that have a value less than 1.
   - **c.** Find the percentage of the data that are not between 0 and 1.

29. Consider measurements taken from a population that has a standard normal distribution.
   - **a.** Find the percentage of the data that have a value between 2 and 3.
   - **b.** Find the percentage of the data that have a value less than 2.
   - **c.** Find the percentage of the data that are not between −2 and 2.

30. Consider measurements taken from a population that has a standard normal distribution.
   - **a.** Find the percentage of the data that have a value between −3 and 1.
   - **b.** Find the percentage of the data that have a value less than −2
   - **c.** Find the percentage of the data that are not between 1 and 3.

Problems 31 through 38

Suppose measurements are taken from a population that has a standard normal distribution. Find the percentage of measurements that are in the specified interval.

31. **a.** between 0 and 1.2.
   b. between −2.3 and 0.
   c. between −0.7 and 1.8.

32. **a.** between 0 and 2.1.
   b. between −1.3 and 0.
   c. between −0.1 and 0.2.

33. **a.** greater than 1.8.
   b. less than 1.8.

34. **a.** greater than −1.3.
   b. less than 1.3.

35. **a.** between 0 and 1 standard deviation above the mean.
   b. between 2 standard deviations below the mean and 0.
   c. between 2 standard deviations below the mean and 1 standard deviation above the mean.

36. **a.** between 0 and 2 standard deviations above the mean.
   b. between 1 standard deviation below the mean and 0.
   c. between 1 standard deviation below the mean and 2 standard deviations above the mean.

37. **a.** within 1 standard deviation of the mean.

b. not within 3 standard deviations of the mean.

38. **a.** within 2 standard deviations of the mean.

b. not within 1 standard deviation of the mean.

39. For measurements taken from a population for which the distribution of measurements can be assumed to be normal, suppose the mean is 16 inches and the standard deviation is 2 inches.

a. What measurement is 1 standard above the mean?

b. What measurement is 2 standard deviations below the mean?

c. What percentage of measurements are within 2 standard deviations of the mean?

d. What percentage of measurements are greater than 16 inches?

40. For measurements taken from a population for which the distribution of measurements can be assumed to be normal, suppose the mean is 327 inches and the standard deviation is 54 inches.

a. What measurement is 2 standard deviations above the mean?

b. What measurement is 1 standard deviation below the mean?

c. What percentage of measurements are within 1 standard deviation of the mean?

d. What percentage of measurements are less than 327 inches?

41. Mensa, founded in 1946, was created as a society for intelligent people. The requirement for membership is a high IQ, which is defined as an IQ in the top 2% of the population. It is known that IQ scores are normally distributed with a mean of 100 and a standard deviation of 15. Approximately what is the lowest IQ accepted for membership into Mensa?

42. The diastolic blood pressure of women between 18 and 70 years of age is normally distributed with a mean of 77 mm Hg and a standard deviation of 10 mm Hg. Suppose a drug company will test a new blood-pressure-lowering drug, and women with high blood pressures are needed for the study. If high blood pressure is defined as having a diastolic blood pressure in the top 7% of the population, approximately what is the lowest diastolic blood pressure that a woman could have and still be included in the study?

43. The serum total cholesterol for women ages 20 to 74 years is normally distributed with a mean of 197 mg/dL and a standard deviation of 43.1 mg/dL. A doctor will conduct a study to see if a high-fat and low-carbohydrate diet will increase levels of serum total cholesterol. For the study, women ages 20 to 74 with normal serum total cholesterol levels will follow the diet for 6 weeks. Suppose a normal serum total cholesterol level is defined as a level that is within 1.5 standard deviations of the mean.

a. What should the serum total cholesterol levels be for women who will be included in this study?

b. Out of 5000 women, approximately how many would be eligible to participate in the study?

44. The serum total cholesterol for men ages 20 to 74 years is normally distributed with a mean of 200 mg/dL and a standard deviation of 44.7 mg/dL. A doctor will conduct a study to see if a low-fat and high-carbohydrate diet will decrease levels of serum total cholesterol. For the study, men ages 20 to 74 with high serum total cholesterol levels will use the diet for 6 weeks. Suppose high serum total cholesterol is defined as a level that is greater than 2.5 standard deviations above the mean.

a. What should the serum total cholesterol levels be for men who will be included in this study?

b. Out of 3500 men, approximately how many would be eligible to participate in the study?

Extended Problems

45. The shape of the normal distribution is completely determined by specifying the mean and the standard deviation. The standard deviation is always a positive quantity and is a measure of the spread of the data. Programs are available on the Internet that allow a user to input a value for the mean and a value for the standard deviation and see the resulting normal distribution. One such program can be found at www.stattucino.com/berrie/dsl/, or use search keywords "normal distribution applet" on the Internet. To sketch a normal distribution, input a mean and a standard deviation and then click the "draw" button. Create several different normal curves by adjusting the values of the mean and standard deviation. Summarize your observations and include sketches.

46. In this section, you learned how to use a table (Table 11.3) to find the area in an interval and below a standard normal curve. There are calculators on the Internet that nicely calculate and illustrate areas under any normal curve between two values. One such site can be found at www.coe.tamu.edu/~strader/Mathematics/Statistics/NormalCurve/ or use search keywords "normal probability calculator." The user may input a mean and a standard deviation and then click and drag the vertical bars to see how the area changes. This program uses blue to show the area under the curve and between two values; it uses red to show the area under the curve and outside the interval. Visit this site to examine the relationship between areas and standard deviation. Consider the following explorations.

a. For the mean $\mu = 8$ and the standard deviation $\sigma = 2$, set the vertical bars to show the area under the curve from $\mu - \sigma = 6$ to $\mu + \sigma = 10$. Make a note of the calculated area.

b. For the mean $\mu = 8$ and the standard deviation $\sigma = 2$, set the vertical bars to show the area under the curve from $\mu - 2\sigma = 4$ to $\mu + 2\sigma = 12$. Make a note of the calculated area.

c. For the mean $\mu = 8$ and the standard deviation $\sigma = 2$, set the vertical bars to show the area under the curve from $\mu - 3\sigma = 2$ to $\mu + 3\sigma = 14$. Make a note of the calculated area.

d. Choose a different mean and standard deviation, and repeat parts (a), (b), and (c). How do the areas compare in each case to the areas you found for the normal curve with a mean of 8 and a standard deviation of 2? Is this a coincidence? Repeat parts (a) to (c) once more, using a different mean and standard deviation. What can you conclude?

47. At the beginning of this chapter, you read about Ronald A. Fisher and W. S. Gossett, who both made significant strides in the field of statistics. Research other famous statisticians, such as Carl Gauss, Karl Pearson, Florence Nightingale, Jerzy Neyman, and Thomas Bayes. What were their important contributions to statistics? Write a report to summarize your findings.

11.2 Applications of Normal Distributions

INITIAL PROBLEM

An automaker is considering two different suppliers of a certain critical engine part. The part must be within 0.012 mm of its required size or the engine will fail. The automaker will carefully measure all parts from the suppliers before installing them. Any parts outside the acceptable range will be discarded. Supplier A charges \$120 for 100 parts and guarantees that the actual part sizes will have a mean equal to the required size and a standard deviation of 0.004 mm. Supplier B charges \$90 for 100 parts and guarantees that the actual part sizes will have a mean equal to the required size and a standard deviation of 0.012 mm. Which supplier should the automaker choose?

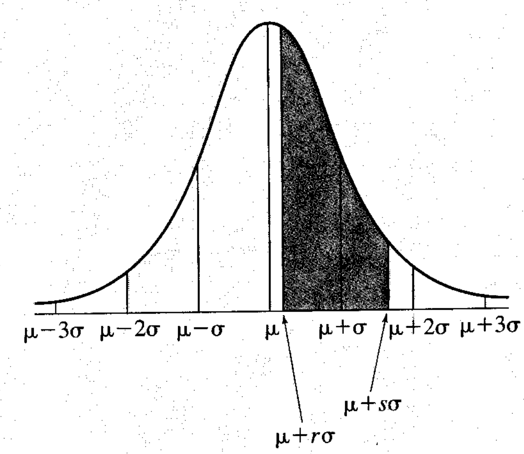
A solution of this Initial Problem is on page 723.

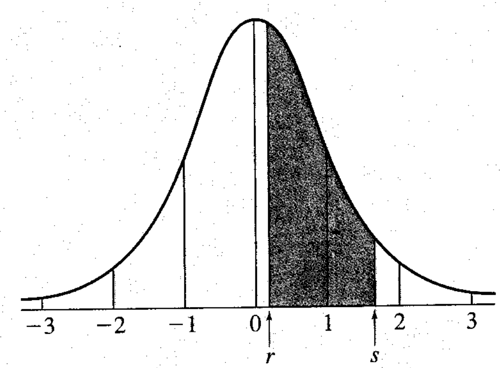
THE RELATIONSHIP AMONG NORMAL DISTRIBUTIONS

We have seen that a critical property of all normal distributions is that if the endpoints of an interval are described by the number of standard deviations that they are above or below the mean, then the percentage of the data in that interval is the same for all normal distributions. In particular, this is true when comparing any normal distribution to the standard normal distribution, which has a mean of 0 and a standard deviation of 1. For example, in Section 11.1, we learned from the standard normal distribution tables that approximately 34% of the data in a standard normal distribution lies between $z = 0$ (the mean) and $z = 1$ (1 standard deviation more than the mean). Similarly, approximately 34% of the data in *any* normal distribution will lie between the mean and the value that is 1 standard deviation more than the mean. We next state this idea more precisely.

RELATIONSHIP BETWEEN NORMAL DISTRIBUTIONS AND THE STANDARD DISTRIBUTION

Suppose a data set is represented by a normal distribution with mean μ and standard deviation σ. The percentage of the data that lies between $\mu + r\sigma$ and $\mu + s\sigma$ is the same as the percentage of the data in a standard normal distribution that lies between r and s.





In Section 11.1, we learned how to use tables to find the percentage of data in a standard normal distribution that lies in a particular interval. The property just stated allows us to find the percentage of data in a particular interval for *any* normal distribution by comparing that distribution to the standard normal distribution.

EXAMPLE 11.7 Approximately 10% of the data in a standard normal distribution are within $\frac{1}{8}$ of a standard deviation of the mean, that is, between $-\frac{1}{8} = -0.125$ and $\frac{1}{8} = +0.125$. Suppose the measurements on a certain population are normally distributed with mean of 112 and standard deviation of 24. What numbers in the latter distribution would correspond to the $-\frac{1}{8}$ and $\frac{1}{8}$?

SOLUTION We are given $\mu = 112$ and $\sigma = 24$. We know that about 10% of the data in a standard normal distribution lie between $-\frac{1}{8}$ and $+\frac{1}{8}$. Therefore, by the property stated above, it must also be true that 10% of the measurements of the population in question lie between the values corresponding to the mean plus $\frac{1}{8}$ of a standard deviation and the mean minus $\frac{1}{8}$ of a standard deviation. In this case, $r = -\frac{1}{8}$ and $s = \frac{1}{8}$ For the nonstandard normal distribution, we have the corresponding values $\mu + r\sigma = 112 + \left(-\frac{1}{8}\right)24 = 109$ and $\mu + s\sigma = 112 + \left(\frac{1}{8}\right)24 = 115$. Because these values are $\frac{1}{8}$ of a standard deviation above and below the mean of 112, we know that 10% of the data will lie between 109 and 115. ■

The next example shows how to apply similar calculations to a measurement taken on a human population.

EXAMPLE 11.8 One factor associated with a person's risk of coronary heart disease is his or her level of HDL cholesterol, the "good" cholesterol. The HDL cholesterol levels for a certain group of women are approximately normally distributed, with a mean of 64 mg/dL and a standard deviation of 15 mg/dL. Determine the percentage of these women who have HDL cholesterol levels between 19 and 109 mg/dL by comparing this normal distribution to the standard normal distribution.

SOLUTION We know that the mean HDL cholesterol level is 64 mg/dL, so this value corresponds to the mean of 0 in the standard normal distribution. A cholesterol level of 109 is $109 - 64$, or 45, more than the mean of 64, and a cholesterol level of 19 is $64 - 19$, or 45, less than the mean. Because the standard deviation is 15, an HDL cholesterol level of 109 mg/dL is 3 standard deviations more than the mean, and an HDL cholesterol level of 19 mg/dL is 3 standard deviations less than the mean. Thus, a cholesterol level of 109 corresponds to a value of 3 in the standard normal distribution, and a cholesterol level of 19 corresponds to a value of -3 in the standard normal distribution.

From Table 11.3, we know that the area under the standard normal curve between $z = 0$ and $z = 3$ is 0.4987. Thus, 49.87% of the data in the standard normal distribution lie between 0 and 3. Also, by the symmetry of the curve, 49.87% of the data in the standard normal distribution lie between $z = -3$ and $z = 0$. Thus, $2 \times 0.4987 = 0.9974 = 99.74\%$ of the data lie between $z = -3$ and $z = 3$. Approximately 99.74% of the women's HDL cholesterol levels, therefore, will lie between 19 and 109 mg/dL. ■

Next we will generalize the result of Example 11.8 by considering the areas of certain regions under any normal distribution curve.

THE 68–95–99.7 RULE FOR NORMAL DISTRIBUTIONS

In Section 11.1, we learned that tables in reference books have information about the standard normal distribution. If these tables are not handy, we can remember three facts about the standard normal distribution. Example 11.8 illustrated one of these facts, namely that about 99.74% of the data in any normal distribution lie within 3 standard deviations of the mean. The following property summarizes three important facts about *all* normal distributions, whether they are standard normal distributions or not.

THE 68–95–99.7 RULE FOR NORMAL DISTRIBUTIONS

Approximately 68% of the measurements in any normal distribution lie within 1 standard deviation of the mean.

Approximately 95% of the measurements in any normal distribution lie within 2 standard deviations of the mean.

Approximately 99.7% of the measurements in any normal distribution lie within 3 standard deviations of the mean.

EXAMPLE 11.9 Suppose a company is concerned about the number of claims of carpal tunnel syndrome filed by employees working in their mail-order department. These employees spend most of their workday using one hand to operate a computer mouse. To reduce the number of injuries, the company commissions a manufacturer to design a new computer mouse to better fit the hands of the employees who will use them. In order to create the mouse, the designers must consider the size of the male and female employees' hands. The designers have learned that lengths of women's hands are normally distributed with a mean of 17 cm and a standard deviation of 1 cm. The designers have tentative plans for a new computer mouse that will comfortably accommodate hands that range from 15 cm to 19 cm in length. What percentage of women have hands with lengths in this range?

SOLUTION Notice that $17 - 15 = 2$, so a woman's hand length of 15 cm is 2 standard deviations below the mean. Similarly, since $19 - 17 = 2$, a woman's hand length of 19 cm is 2 standard deviations above the mean. So, we must find the percentage of women who have hands that are within 2 standard deviations of the mean length. By the 68–95–99.7 rule, approximately 95% of the women have hand lengths that are between 15 cm and 19 cm. ■

We next examine what the 68–95–99.7 rule means in terms of the graph of a normal distribution. Figure 11.15 gives a visual summary of the 68–95–99.7 rule.

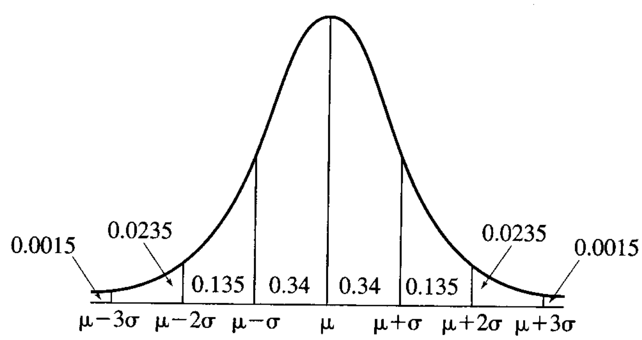


Figure 11.15

By using the symmetry of the normal distribution and the fact that the total area under the curve is 1, we can use the 68–95–99.7 rule to compute the areas of all the regions shown in the figure. The figure shows that the area between the point that is 1 standard deviation below the mean ($\mu - \sigma$) and the point that is 1 standard deviation above the mean ($\mu + \sigma$) is $0.34 + 0.34$, or 68% of the data. Likewise, the area between $\mu - 2\sigma$ and $\mu + 2\sigma$ is $0.135 + 0.34 + 0.34 + 0.135$, or 95% of the data. Finally, the area between $\mu - 3\sigma$ and $\mu + 3\sigma$ is $0.0235 + 0.135 + 0.34 + 0.34 + 0.135 + 0.0235 = .0997$, or 99.7% of the data.

The next example shows how to combine the areas shown in Figure 11.15 to determine a desired area.

EXAMPLE 11.10 The lake sturgeon is a large fish native to the Great Lakes. In 2001, the lake sturgeon caught by fishermen in Saginaw Bay of Lake Huron had a mean length of about 114 cm and a standard deviation of approximately 29 cm. (*Source:* Midwest.fws.gov/alpena/rpt-sagbay01.pdf.) If the lengths of these fish had a normal distribution, determine the following.

a. What percentage of lake sturgeon caught had lengths between 56 cm and 143 cm?

b. What percentage of lake sturgeon caught were not between 56 cm and 143 cm in length?

SOLUTION

a. Notice that $114 - 56 = 58$, which is 2×29, or twice the standard deviation of 29. Therefore, a length of 56 cm is 2 standard deviations below the mean. Also, because $143 - 114 = 29$, which equals the standard deviation, a length of 143 is 1 standard deviation above the mean. Thus, the data in question lie between

$\mu - 2\sigma$ and $\mu + \sigma$. Figure 11.15 shows three regions between $\mu - 2\sigma$ and $\mu + \sigma$. The sum of their areas is $0.135 + 0.34 + 0.34 = 0.815$. Therefore, 81.5% of the lake sturgeon caught were between 56 and 143 cm long.

Figure 11.16 illustrates the solution to Example 11.10(a).

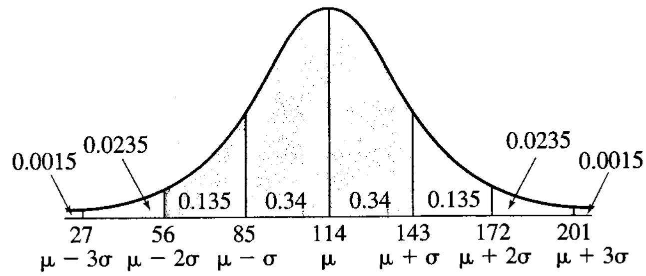


Figure 11.16

b. To find the percentage of measurements that do *not* lie between 56 cm and 143 cm, remember that the total area beneath the curve is 100%. Subtracting 81.5% from 100%, we estimate that 100% − 81.5, or 18.5%, of the lake sturgeon caught in Saginaw Bay had lengths that were either less than 56 cm or greater than 143 cm. Verify that 18.5% is the correct percentage by adding the unshaded areas in Figure 11.16. ■

We have so far been able to find areas under a normal curve by relating points in the normal distribution to corresponding points in a standard normal distribution. In the discussion which follows, we will formalize this process.

POPULATION *z*-SCORES

We can convert any normal distribution of data to a standard normal distribution simply by changing to a new measurement scale called the "population z-score." This scale corresponds to the standard normal distribution, which has its mean at the origin and the standard deviation as its unit. Each data point in a normal distribution is assigned a z-score based on its position relative to the mean of the distribution. For example, if a data point is 1 standard deviation greater than the mean of a distribution, it is assigned a z-score of 1. If a data point has a z-score of -2.5, then it is 2.5 standard deviations less than the mean of the distribution. In Example 11.10, we found that the percentage of data points that lie between data points corresponding to $z = -2$ and $z = 1$ is 81.5%. The formal definition of z-score is given next.

> ***Definition***
>
> **POPULATION *z*-SCORE**
>
> The **population *z*-score** of a measurement, x, is given by
>
> $$z = \frac{x - \mu}{\sigma},$$
>
> where μ is the population mean and σ is the population standard deviation.

Notice that $|z|$ is the number of standard deviations that a data point x is away from the mean. Also, if $x < \mu$, the value of z is negative, which means it is to the left of zero, so the corresponding data point is less than the mean of the distribution. On the other hand, if $x > \mu$, the value of z is positive and to the right of zero, so the corresponding data point lies to the right of the mean. Figure 11.17 illustrates this idea. Each point on the horizontal axis of Figure 11.15 represents a z-score. For example, we may represent the point $\mu + \sigma$ as $z = 1$, and so on.

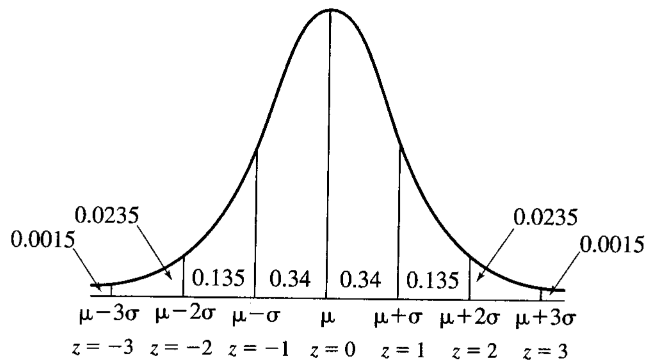


Figure 11.17

EXAMPLE 11.11 Suppose a certain normal distribution has mean 4 and standard deviation 3. Compute the z-scores of the measurements -1, 2, 3, 5, and 9 in the standard normal distribution.

SOLUTION Table 11.4 lists the z-scores, which were calculated using the definition of a z-score with $\mu = 4$ and $\sigma = 3$.

Table 11.4

Measurement	z-score
$x = -1$	$\frac{x - \mu}{\sigma} = \frac{-1 - 4}{3} \approx -1.67$
$x = 2$	$\frac{x - \mu}{\sigma} = \frac{2 - 4}{3} \approx -0.67$
$x = 3$	$\frac{x - \mu}{\sigma} = \frac{3 - 4}{3} \approx -0.33$
$x = 5$	$\frac{x - \mu}{\sigma} = \frac{5 - 4}{3} \approx 0.33$
$x = 9$	$\frac{x - \mu}{\sigma} = \frac{9 - 4}{3} \approx 1.67$

■

A z-score is a measure of relative standing. Measurements that are below the mean have negative z-scores and those that are above the mean have positive z-scores. A z-score of 0 indicates that the measurement is at the population mean, and z-scores with small absolute values indicate measurements that are near the mean. Figure 11.18 shows the relationship between the distribution in Example 11.11 (x-values) and the standard normal distribution (z-scores).

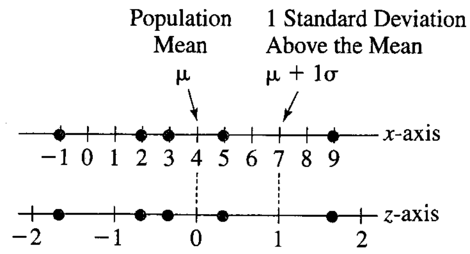


Figure 11.18

Each dot on the x-axis represents an x-value from Example 11.11, and each dot on the z-axis represents a corresponding z-score. Note that the z-score describes the location of the corresponding measurement along a new z-axis that has its origin at the population mean and the population standard deviation as its unit length.

COMPUTING WITH NORMAL DISTRIBUTIONS

For a set of measurements with a normal distribution, replacing each measurement with its z-score gives a standard normal distribution. We can then use Table 11.1 or Table 11.3 to determine the percentage of measurements in particular intervals. For example, suppose we wish to compute the percentage of people with an IQ between 100 and 109. It is known that IQ scores are normally distributed with a mean of 100 and a standard deviation of 15. Thus, an IQ of 100 corresponds to a z-score of 0 and an IQ of 109 corresponds to a z-score of $z = \frac{x - \mu}{\sigma} = \frac{109 - 100}{15} = 0.6$. From Table 11.3, we see that the area under the standard normal distribution curve from 0 to 0.6 is 0.2257. We conclude that 22.57% of z-scores lie in that interval, and thus 22.57% of IQ scores are between 100 and 109. (See Figure 11.19.)

Tidbit

Intelligence testing originated in France, when, in 1904, psychologist Alfred Binet was commissioned by the French government to find a method to differentiate between children who were intellectually normal and those who needed special attention. His work, along with that of Theophile Simon, led to the development of the Simon-Binet scale. On this scale, a score of 100 is considered "average." A score of 115 is considered "high average," and a score of 130 or more is considered "gifted."

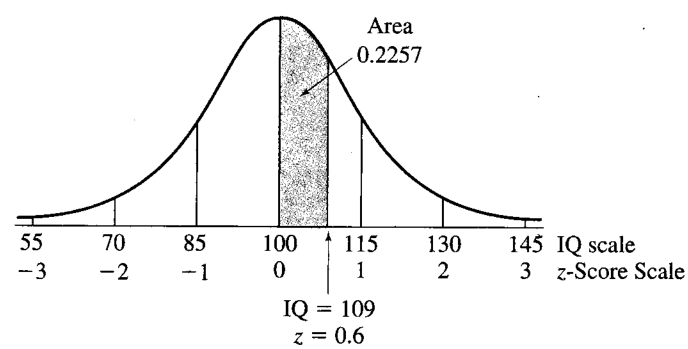


Figure 11.19

Note that the horizontal axis on the graph indicates both IQ scores and the corresponding z-scores.

EXAMPLE 11.12 The New York City Marathon has been held on the first Sunday in November every year since 1970, and the event now attracts about 30,000 runners. In 1996, the distribution of finishing times for the 26.2-mile race was approximately normal, with a mean of about 260 minutes and a standard deviation of about 50 minutes. What percentage of the finishers in 1996 had times that were between 285 minutes and 335 minutes?

SOLUTION A finishing time of 285 minutes corresponds to a z-score of $\frac{x-\mu}{\sigma} = \frac{285-260}{50} = \frac{25}{50} = 0.5$. A finishing time of 335 minutes corresponds to a z-score of $\frac{x-\mu}{\sigma} = \frac{335-260}{50} = \frac{75}{50} = 1.5$. From Table 11.3, we see that the area under the curve from 0 to 0.5 in a standard normal distribution is 0.1915. Similarly, the area from 0 to 1.5 is 0.4332. The area under the standard normal distribution curve from 0.5 to 1.5 is the difference of these two areas, which is $0.4332 - 0.1915$, or 0.2417, as illustrated in Figure 11.20.

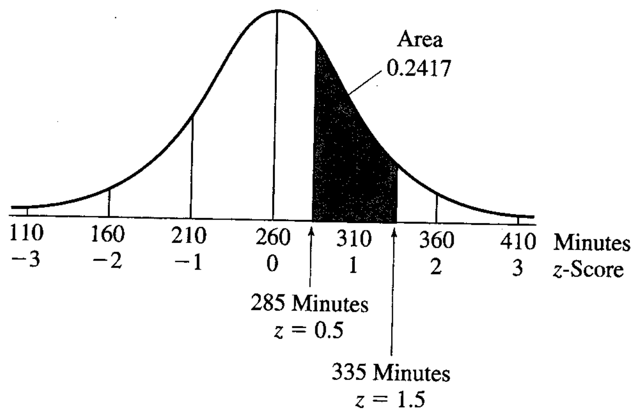


Figure 11.20

We may conclude, therefore, that 24.17% of the finishing times were between 285 and 335 minutes. ■

In Example 11.8, we examined the distribution of HDL cholesterol levels in a group of women. In the next example, we take another look at that same group of women.

EXAMPLE 11.13 Recall that the HDL cholesterol levels for a certain group of women were approximately normally distributed with a mean of 64 mg/dL and a standard deviation of 15 mg/dL. If an HDL cholesterol level less than 40 signals that a woman is "at increased risk" for coronary heart disease, what percentage of the women in this group are at increased risk?

SOLUTION We must first determine what point in the standard normal distribution corresponds to an HDL cholesterol level of 40 mg/dL. We find that the corresponding z-score is $z = \frac{x-\mu}{\sigma} = \frac{40-64}{15} = -1.6$. The area under the standard normal curve to the left of $z = -1.6$ is the same as the area under the standard normal curve to the right of $z = 1.6$. Using Table 11.3, we find that the area between $z = 0$ and $z = 1.6$ is 0.4452. To find the area to the right of $z = 1.6$, we use the fact that the area under the curve to the right of the mean is 0.5. Thus, the area to the right of $z = 1.6$ is $0.5 - 0.4452 = 0.0548$. In this particular group of women, approximately 5.48% of the women are at "increased risk" for coronary heart disease on the basis of their HDL cholesterol levels. ■

SOLUTION OF THE INITIAL PROBLEM

An automaker is considering two different suppliers of a certain critical engine part. The part must be within 0.012 mm of its required size or the engine will fail. The automaker will carefully measure all parts from the suppliers before installing them. Any parts outside the acceptable range will be discarded. Supplier A charges \$120 for 100 parts and guarantees that the actual part sizes will have a mean equal to the required size and a standard deviation of 0.004 mm. Supplier B charges \$90 for 100 parts and guarantees that the actual part sizes will have a mean equal to the required size and a standard deviation of 0.012 mm. Which supplier should the automaker choose?

SOLUTION At first glance, supplier B looks like the best choice because the parts from this supplier are less expensive. Notice that parts from supplier B cost $\frac{\$90}{100}$, or \$0.90 each, while parts from supplier A cost $\frac{\$120}{100}$, or \$1.20 each.

However, let's take a closer look. We will consider how many parts from each supplier are acceptable and what each acceptable part costs. We assume that in each case the part sizes are normally distributed. For supplier A, the tolerance of 0.012 mm is 3 times the standard deviation of 0.004. Thus, all parts from supplier A that are within 3 standard deviations of the mean will be usable. By the 68–95–99.7 rule, we know that 99.7 of every 100 parts (on average) will be within 3 standard deviations of the mean and will be acceptable. So, the average cost of each acceptable part from supplier A is $\frac{\$120}{99.7}$, or about \$1.20.

On the other hand, for supplier B, the tolerance of 0.012 is exactly equal to the standard deviation. Thus, only parts from supplier B that are within 1 standard deviation of the mean will be usable by the automaker. Again, by the 68–95–99.7 rule, 68 of every 100 parts (on average) will be within 1 standard deviation of the mean and will be acceptable. So, the average cost of each acceptable part from supplier B is $\frac{\$90}{68}$, or about \$1.32. The average cost of the acceptable parts from supplier A is less than the average cost of acceptable parts from supplier B. In addition, fewer parts from supplier A will be discarded. Therefore, the automaker should choose supplier A.

PROBLEM SET 11.2

1. A data set is represented by a normal distribution with a mean of 25 and a standard deviation of 3. For each of the following data values, how many standard deviations above or below the mean is it?

   a. 28 **b.** 31 **c.** 22
   d. 26.5 **e.** 20.5 **f.** 16

2. A data set is represented by a normal distribution with a mean of 206 and a standard deviation of 22. For each of the following data values, how many standard deviations above or below the mean is it?

   a. 184 **b.** 272 **c.** 239
   d. 217 **e.** 140 **f.** 206

3. Suppose a data set is represented by a normal distribution with a mean of 52 and a standard deviation of 10.

   a. What data value is 3 standard deviations above the mean?
   b. What data value is 2 standard deviations below the mean?
   c. What data value is 1.5 standard deviations below the mean?
   d. What data value is 2.5 standard deviations above the mean?
   e. What data value is $\frac{1}{4}$ of a standard deviation above the mean?

4. Suppose a data set is represented by a normal distribution with a mean of 125 and a standard deviation of 7.

a. What data value is 2 standard deviations above the mean?

b. What data value is 3 standard deviations below the mean?

c. What data value is 1.5 standard deviations below the mean?

d. What data value is 2.5 standard deviations above the mean?

e. What data value is $\frac{1}{5}$ of a standard deviation below the mean?

5. Approximately 20% of the data in a standard normal distribution are between $-\frac{1}{4}$ and $\frac{1}{4}$, or within $\frac{1}{4}$ of a standard deviation of the mean. Suppose the measurements on a population are normally distributed with mean 84 and standard deviation 8.

a. What data value is $\frac{1}{4}$ of a standard deviation above the mean?

b. What data value is $\frac{1}{4}$ of a standard deviation below the mean?

c. What percentage of the measurements in the population lie between 82 and 86?

6. Approximately 50% of the data in a standard normal distribution are between $-\frac{2}{3}$ and $\frac{2}{3}$, or within $\frac{2}{3}$ of a standard deviation of the mean. Suppose the measurements on a population are normally distributed with mean 145 and standard deviation 12.

a. What data value is $\frac{2}{3}$ of a standard deviation above the mean?

b. What data value is $\frac{2}{3}$ of a standard deviation below the mean?

c. What percentage of the measurements of the population lie between 137 and 153?

7. Suppose that turkeys from a certain ranch have weights that are normally distributed with a mean of 12 pounds and a standard deviation of 2.5 pounds. Use the 68–95–99.7 rule.

a. What percentage of turkeys have weights between 9.5 pounds and 14.5 pounds?

b. What percentage of turkeys have weights between 4.5 pounds and 19.5 pounds?

c. What percentage of turkeys have weights between 7 pounds and 12 pounds?

8. Suppose that heights of a certain type of tree are normally distributed with a mean of 154 feet and a standard deviation of 8.2 feet. Use the 68–95–99.7 rule.

a. What percentage of trees have heights between 129.4 feet and 178.6 feet?

b. What percentage of trees have heights between 145.8 feet and 162.2 feet?

c. What percentage of trees have heights between 154 feet and 170.4 feet?

9. A certain population has measurements that are normally distributed with a mean of μ and a standard deviation of σ. Use the 68–95–99.7 rule.

a. Find the percentage of measurements that are between $\mu - \sigma$ and $\mu + 2\sigma$.

b. Find the percentage of measurements that are between $\mu - 3\sigma$ and $\mu + \sigma$.

c. Find the percentage of measurements that are not between $\mu - \sigma$ and $\mu + \sigma$.

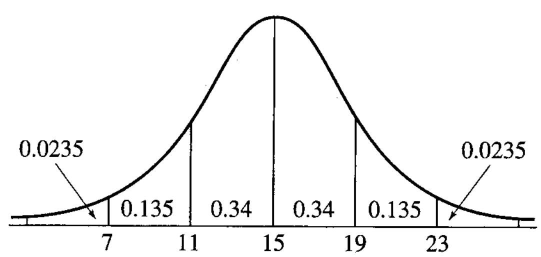
10. A certain population has measurements that are normally distributed with a mean of μ and a standard deviation of σ. Use the 68–95–99.7 rule.

a. Find the percentage of measurements that are between $\mu - 2\sigma$ and $\mu + 2\sigma$.

b. Find the percentage of measurements that are between $\mu - 3\sigma$ and $\mu + 2\sigma$.

c. Find the percentage of measurements that are not between $\mu - 3\sigma$ and $\mu + \sigma$.

11. Consider the following normal distribution.



a. What is the mean of the population? How do you know?

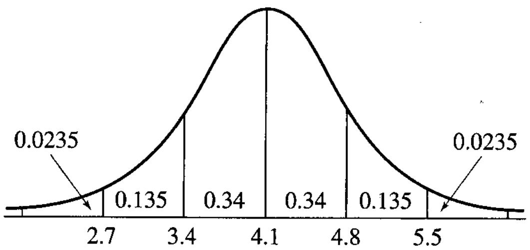
b. What is the standard deviation of the population? How do you know?

c. What percentage of the measurements are between 11 and 23?

d. What percentage of the measurements are between 7 and 19?

e. What percentage of the measurements are not between 7 and 23?

12. Consider the following normal distribution.



a. What is the mean of the population? How do you know?

b. What is the standard deviation of the population? How do you know?

c. What percentage of the measurements are between 4.1 and 5.5?

d. What percentage of the measurements are between 2.7 and 4.8?

e. What percentage of the measurements are not between 3.4 and 5.5?

13. Suppose a normal distribution has mean 10 and standard deviation 2. Find the z-scores of the measurements 9, 10, 11, 14, and 17.

14. Suppose a normal distribution has mean 20.5 and standard deviation 0.4. Find the z-scores of the measurements 19.3, 20.2, 20.5, 21.3, and 23.

15. Recall that IQ scores are normally distributed with mean 100 and standard deviation 15. Find the z-scores of the IQ scores 64, 80, 96, 111, 136, and 145.

16. Suppose that the weights of rabbits have a normal distribution with a mean of 7.1 pounds and a standard deviation of 0.6 pounds. Find the z-scores of the weights 5.9, 6.2, 7.1, 7.2, 8.4, and 9 pounds.

17. Suppose that the weights of checked luggage for individuals checking in at a particular airport have a normal distribution of 55.6 pounds and a standard deviation of 11.3 pounds. Find the z-scores of weights 45.16, 49.82, 55.20, and 58.63 pounds.

18. Suppose that insect lifetimes for a particular species have a normal distribution with a mean of 812 hours and a standard deviation of 19 hours. Find the z-scores for lifetimes of 749, 766, 791, 801, 833, and 842 hours.

19. In a normally distributed data set, find the value of the mean if the following additional information is given.

a. The standard deviation is 4.25 and the z-score for a data value of 52.1 is 1.9.

b. The standard deviation is 0.6 and the z-score for a data value of 2 is -2.3.

20. In a normally distributed data set, find the value of the mean if the following additional information is given.

a. The standard deviation is 3.5 and the z-score for a data value of 15.3 is 1.2.

b. The standard deviation is 0.8 and the z-score for a data value of 4.9 is -0.6.

21. In a normally distributed data set, find the value of the standard deviation if the following additional information is given.

a. The mean is 9.8 and the z-score for a data value of 10.3 is 2.

b. The mean is 577 and the z-score for a data value of 533 is -0.5.

22. In a normally distributed data set, find the value of the standard deviation if the following additional information is given.

a. The mean is 226.2 and the z-score for a data value of 230 is 0.2.

b. The mean is 14.6 and the z-score for a data value of 5 is -0.3.

23. Suppose that there are 100 franchises of Betty's Boutique in similar shopping malls across America. The gross Saturday sales of these boutiques are approximately normally distributed with a mean of \$4610 and a standard deviation of \$370.

a. Find the z-scores of each of the following gross Saturday sales amounts: \$3870, \$4425, and \$5535.

b. What percentage of the Betty's Boutique franchises had gross Saturday sales between \$4425 and \$5535? Use the z-scores you found in part (a) and Table 11.3.

c. What percentage had gross Saturday sales between \$3870 and \$5535? Use the z-scores you found in part (a) and Table 11.3.

d. What percentage of stores had gross Saturday sales less than \$5535?

24. The lifetime of a certain brand of passenger tire is approximately normally distributed with a mean of 41,500 miles and a standard deviation of 1950 miles.

a. Find the z-scores of each of the following tire lifetimes: 38,575; 41,500; and 46,765.

b. What percentage of this brand of tires will have lifetimes between 38,575 and 41,500 miles? Use the z-scores you found in part (a) and Table 11.3.

c. What percentage of tires will have lifetimes between 38,575 and 46,765 miles? Use the z-scores you found in part (a) and Table 11.3.

d. What percentage of tires will have lifetimes of more than 46,765 miles?

25. Suppose that a certain breed of dog has a mean weight of 11 pounds with a standard deviation of 3.5 pounds. Also, suppose that the weights of this breed of dog are approximately normally distributed.

a. What percentage of dogs will weigh between 5.75 and 16.25 pounds?

b. What percentage of dogs will weigh between 11 and 18 pounds?

c. What percentage of dogs will weigh more than 15.55?

d. What percentage of dogs will weigh less than 1.2 pounds?

26. Suppose that a certain insect has a mean lifespan of 5.6 days with a standard deviation of 1.2 days. Assume the lifespan of this insect is approximately normally distributed.

a. Calculate the percentage of insects with a lifespan between 3.2 and 8 days.

b. Calculate the percentage of insects with a lifespan between 2.6 and 8.6 days.

c. What percentage of insects will live longer than 3.32 days?

d. What percentage of insects will live less than 6.56 days?

Problems 27 through 32

The SAT is a standardized, 3-hour test designed to measure verbal and mathematical abilities of college-bound seniors. Many colleges and universities use the scores from the SAT as part of their admissions processes because the result of the test is one predictor of how well incoming students may do in college. Until 2005, students could earn at most 1600 points on the SAT: 800 for the verbal section and 800 for the math section. The scores for the SAT are normally distributed. In the year 2003, a total of 1,406,324 high school students took the SAT. (*Source:* www.collegeboard.com.)

27. For the 2003 SAT scores in the math section only, the mean was 507 and the standard deviation was 111.

a. What percentage of students had a math score between 300 and 500 points? Round your z-scores to the nearest tenth.

b. What percentage of students had a math score less than 200 points? Round your z-score to the nearest tenth.

c. Approximately how many students had a math score of at least 400? Round your z-score to the nearest tenth.

28. For the 2003 SAT scores in the verbal section only, the mean was 519 and the standard deviation was 115.

a. What percentage of students had a verbal score between 300 and 500 points? Round your z-scores to the nearest tenth.

b. What percentage of students had a verbal score less than 200 points? Round your z-score to the nearest tenth.

c. Approximately how many students had a verbal score of at least 400? Round your z-score to the nearest tenth.

29. Boston College considers a variety of factors when deciding which students to admit. In 2003, a student was considered "competitive" in the application process if he or she had both a math SAT score and a verbal SAT score in the mid to high 600s. Refer to problems 27 and 28 for mean and standard deviation information. Round your z-score to the nearest tenth.

a. What percentage of students had a math score of at least 650?

b. What percentage of students had a verbal score of at least 650?

30. Harvard University considers a variety of factors, including SAT scores, when deciding which students to admit to the freshman class. Harvard does not have any cutoff SAT scores, but they consider a student "competitive" in the application process if he or she had math scores as well as verbal score in the 700s. Refer to problems 27 and 28 for mean and standard deviation information. Round your z-scores to the nearest tenth.

a. What percentage of students had a math SAT score of at least 700?

b. What percentage of students had a verbal SAT score of at least 700?

31. Of the students whose parents did not earn a high-school diploma, the mean verbal score on the 2003 SAT was 413 with a standard deviation of 100, and the mean math score on the 2003 SAT was 443 with a standard deviation of 114.

a. What percentage of these students had a verbal score between 500 and 600? Round your z-scores to the nearest tenth.

b. What percentage of these students had a math score between 500 and 600? Round your z-scores to the nearest tenth.

c. What percentage of these students had a verbal score less than 300 points? Round your z-score to the nearest tenth.

d. What percentage of these students had a math score less than 300 points? Round your z-score to the nearest tenth.

32. Of the students whose parents have a graduate degree, the mean verbal score on the 2003 SAT was 559 with a standard deviation of 107, and the mean math score on the 2003 SAT was 569 with a standard deviation of 111.

a. What percentage of these students had a verbal score between 500 and 600? Round your z-scores to the nearest tenth.

b. What percentage of these students had a math score between 500 and 600? Round your z-scores to the nearest tenth.

c. What percentage of these students had a verbal score less than 300 points? Round your z-score to the nearest tenth.

d. What percentage of these students had a math score less than 300 points? Round your z-score to the nearest tenth.

33. Assume the distribution of the annual total rainfall in Guatemala is approximately normal with a mean of 955 mm and a standard deviation of 257 mm. In 2001, a 4-month drought during the rainy season damaged crops and caused 65,000 people to suffer life-threatening malnutrition. In what percentage of years would Guatemala suffer from drought conditions, that is, receive less than 600 mm of rain? Round your z-score to the nearest tenth.

34. A study recorded the serum total cholesterol levels for 7429 females ages 20 to 74. The study found that the serum total cholesterol was approximately normally distributed, with a mean of 204 mg/dL and a standard deviation of 44.2 mg/dL. According to the American Heart Association, high total serum cholesterol is 240 mg/dL or above. Estimate the percentage of females in this study who had a high total serum cholesterol level.

35. A study partially funded by Novartis Animal Health monitored the number of fleas when flea-infested dogs and cats were treated monthly with imidacloprid. Scientists counted fleas in the animal's surroundings and on the animal. The following table summarizes the counts of fleas on day 0, 14, and 28, where μ = the mean number of fleas observed. Assume flea counts are approximately normally distributed.

Flea Counts	Day 0	Day 14	Day 28
On animal	$\mu = 41.3$ $\sigma = 58.8$	$\mu = 1.4$ $\sigma = 2.5$	$\mu = 1.7$ $\sigma = 4.7$
In surroundings	$\mu = 13.8$ $\sigma = 11$	$\mu = 12.7$ $\sigma = 14$	$\mu = 1$ $\sigma = 1.7$

(Source: www.vet.ksu.edu/depts/dmp/personnel/faculty/docs/imidfip.doc.)

a. Compare the percentages of animals that had at most one flea observed on them on day 14 and on day 28. Round your z-scores to the nearest tenth. What can you conclude?

b. Compare the percentages of animals that had at most one flea observed in their surroundings on day 14 and on day 28. Round your z-scores to the nearest tenth. What can you conclude?

c. If you treat 50 dogs with imidacloprid, and you count fleas on the animals 28 days later, approximately how many dogs might you find with fewer than 5 fleas? At least 11 fleas?

36. Housefly wing-length measurements are approximately normally distributed with mean length 4.55 mm and standard deviation 0.392 mm. (*Source:* www.seattlecentral.edu.)

a. What percentage of housefly wings are longer than 5 mm? Round your z-scores to the nearest tenth.

b. What percentage of housefly wings are between 3 mm and 4 mm? Round your z-scores to the nearest tenth.

c. If you swat 30 houseflies and measure their wing lengths, approximately how many houseflies might you find with wing lengths between 3.77 mm and 4.16 mm? Between 4.16 mm and 4.55 mm?

Extended Problems

Problems 37 and 38

We can use the normal distribution to estimate the number of measurements in a sample less than a given number. For example, if 16% of measurements from a normal distribution are less than 200, we can expect about 16 measurements out of 100 to be less than 200. Similarly, we can expect about 160 measurements out of 1000 to be less than 200.

37. In problem 24, we looked at mileage ratings for automobile tires. Suppose that the tire company guarantees tires to last at least 38,000 miles and will replace any tire that does not last this long (on a properly aligned automobile). What percentage of tires will have to be replaced? If the cost of such a replacement is \$86, how much will the company expect to pay on each lot of 1000 tires?

38. Suppose a snack food manufacturer claims their boxes of crackers are filled to a mean weight of 1.3 pounds with a standard deviation of 0.2 pound. In an ad campaign, they promise to reimburse customers if the actual weight of the box of crackers is less than 1 pound. If a box of crackers costs $2.50 and there are 1 million boxes sold in a year, what is the expected cost of such a program to the manufacturer?

39. Graphics calculators can graph normal distributions. The equation for the normal distribution with mean μ, standard deviation σ, and variable x is as follows.

$$y = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

a. If $\mu = 0$ and $\sigma = 1$, the equation represents the standard normal curve. Substitute the values $\mu = 0$ and $\sigma = 1$ into the equation and simplify. Graph the equation on your calculator. Think about your window settings. What values of x will make sense? Sketch and label this graph.

b. Keep the graph of the standard normal curve on the screen and graph two more normal curves. For these, continue to use $\mu = 0$, but select different values for σ in each equation. You may have to adjust your window settings to see all three graphs. Sketch and label the three graphs. What can you conclude?

c. Keep the graph of the standard normal curve from part (a) on the screen and delete the other two graphs. Graph two more normal curves: use $\sigma = 1$ and select different values for μ in each equation. You may have to adjust your window settings. Sketch and label the three graphs. What can you conclude?

d. Keep the graph of the standard normal curve from part (a) on the screen and delete the other two graphs. Change both the mean and standard deviation, predict how the new normal curve will compare to the standard normal curve, and graph the new normal curve. Was your prediction correct?

40. Fourteen teams were in the 2003–2004 Women's National Basketball Association (WNBA). Create a frequency histogram of heights for all of the players in the WNBA for the current season and determine if the distribution is approximately normal. Find the mean and the standard deviation using the statistical features of your calculator. Summarize the heights using the 68–95–99.7 rule. On the Internet go to http://sports.espn.go.com/wnba/teams for a list of women on each team's roster and their heights.

41. Research the history of the normal distribution. Who is credited with the "discovery" of the normal distribution? What were some of its earliest applications? Summarize your findings in a report.

42. Cereal and potato chip manufacturers want consumers to be happy with their products. They specify the weight of the contents on the label. However, packaging processes fill cereal boxes and potato chip bags imperfectly. If a consumer opened a bag of potato chips to find only 9 ounces instead of the stated 13.25 ounces, he or she would not be very happy. How much variation in weight is there from bag to bag? Contact, by phone or e-mail, manufacturers of two of the packaged products you usually buy. Ask them for statistical data related to packaging, specifically the mean weight of the contents and the standard deviation. Then calculate the percent of packages that would contain less than 90%, less than 80%, less than 70%, and less than 60% of the stated weight. How many packages of the product do you buy in a year? Suppose, for example, that you buy approximately 40 of the 13.25-ounce packages in a year. Determine how many bags will likely contain less than 12 ounces and how much you are overpaying for the product. Summarize your findings in a table and write a report.

43. We have seen that we can use tables to determine areas under the standard normal curve. Many Internet sites will also calculate percentages related to a normal distribution. Search the Internet to find a site that will allow you to input the mean, standard deviation, and endpoints of an interval and that will return the percentage of values in that interval under the normal curve. Many sites will also graph the information for you. Search the Internet using keywords "normal distribution probability applets." Use the applet to obtain more exact solutions to problems 33 through 36.

11.3 Confidence Intervals and Reliable Estimation

INITIAL PROBLEM

A candy bar company has a promotion in which letters printed inside some of the wrappers earn the buyer a prize. Suppose that you buy 400 of these candy bars and you discover that 25 wrappers are winners. You are thinking of buying another 1000 candy bars. How many wrappers would you expect to have letters printed on the inside?

A solution of this Initial Problem is on page 738.

In many statistical applications, the statement that a population is normally distributed really means that the distribution is "almost" normal. The advantage of assuming that a population is normally distributed is that we can use the standard deviation and mean of the distribution together with the standard normal distribution to analyze the distribution. For example, as we saw in Section 11.2, we can estimate percentages of the population that lie in particular intervals. In this section we investigate a distribution that is known to be almost normal, the population of sample proportions. Knowing how sample proportions are distributed, together with information about their standard deviations, will be the basis for drawing conclusions about populations and assigning a level of confidence to our results.

Tidbit

John F. Kennedy was the first U.S. presidential candidate to strategically use polling during a presidential campaign. He hired Louis Harris to poll each state during the 1959–1960 campaign. It was estimated from one poll that 30% of families were sending children to college but 80% of families were *hoping* to send children to college in the future. As a result, Kennedy focused on education and the public supported him, in part because of his emphasis on education.

SAMPLE PROPORTION

Suppose a poll tells you that 48% of registered voters in the United States support the budget submitted to Congress by the President. The poll in question is based on 413 interviews, and the margin of error is reported to be 5%. How can you interpret these statements? In particular, how can interviews with only 413 people out of 130,000,000 registered voters give reliable information?

In order to answer these questions, we must first introduce the concept of a population proportion. Suppose that in fact 50% of U.S. registered voters support the budget submitted by the President. Because 50% of 130,000,000 is 65,000,000, this means that roughly 65,000,000 registered voters support the budget. The proportion $\frac{65,000,000}{130,000,000} = 50\%$ is called a **population proportion**, since it represents a certain fraction of the entire population under consideration. A population proportion is represented by the letter p. So, for this example, we have $p = 0.50$.

The opinion poll just mentioned had a similar proportion of people who supported the budget. In the sample of 413 people, the pollster divided the number of persons who said they support the budget, 198 in this case, by the total number of persons polled, 413, to arrive at the proportion $\frac{198}{413} \approx 48\%$. This number is called a sample proportion because it represents a certain fraction of the sample. A sample proportion is represented by the symbol $\hat{p}$, read as "p hat." For the presidential budget example, we have $\hat{p} = 0.48$. Notice that the sample proportion (0.48) is not very different from the population proportion (0.50). If the selected sample is representative of the population, the sample proportion will be close to the population proportion.

For a sample of size n, we compute the sample proportion as follows.

> ***Definition***
>
> SAMPLE PROPORTIONS
>
> If a sample of size n is selected from a population, then the fraction of the sample that belongs to a particular group is called the **sample proportion** and is given by
>
> $$\hat{p} = \frac{\text{number in the sample that belong to the group}}{n}.$$

EXAMPLE 11.14 Suppose that 3520 freshmen attend Friendly State College, and 1056 of those freshmen have consumed an alcoholic beverage within the past 30 days. The instructor in a freshman health class of 50 students asks her class to fill out an anonymous survey in which one of the questions asks "Have you consumed an alcoholic beverage within the past 30 days?" Eleven of the students in the class respond "Yes" and the other 39 respond "No." In this situation, what are the population, the sample, the population proportion, and the sample proportion?

SOLUTION The population is the 3520 freshmen students attending Friendly State College. The sample is the set of 50 freshmen in this particular health class, so $n = 50$. We calculate the population proportion as follows:

$$p = \frac{\text{number of students in the population who consumed alcohol in the past 30 days}}{\text{number of students in the population}}$$

$$p = \frac{1056}{3520} = 0.30 = 30\%.$$

We calculate the sample proportion in a similar way:

$$\hat{p} = \frac{\text{number of students in the sample who consumed alcohol in the past 30 days}}{\text{number of students in the sample}}$$

$$\hat{p} = \frac{11}{50} = 0.22 = 22\%.$$ ■

DISTRIBUTION OF SAMPLE PROPORTIONS

In Example 11.14, the sample proportion differs noticeably from the population proportion. The discrepancy could have several explanations. The students in the health class may be more health-conscious and consequently less likely to consume alcohol than typical freshmen, or students in the class may be reluctant to answer the question honestly. For whatever reason, the sample is probably biased; that is, it is probably *not* representative of the entire freshman class.

The sample of 50 freshmen in the health class is just one of many possible samples of size 50 from the college campus. The instructor could have asked the same question of many other groups of 50 students: 50 students living in a particular dormitory, 50 students walking to and from class, 50 students eating in the cafeteria, and so on. It turns out that over 10^{112} different samples of size 50 can be selected from a population of 3520. For each of those 10^{112} different samples, we could calculate the sample proportion. In some cases the calculated sample proportion would be very close to the population proportion of $p = 0.30$, while in other cases it would be very different.

A histogram of the various sample proportions for all possible samples of size 50 from a population of 3520 is shown in Figure 11.21. In the histogram, the population proportion is assumed to be 30%, as in Example 11.14.

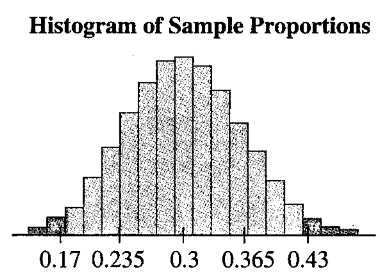


Figure 11.21

Notice that we can closely approximate the histogram in Figure 11.21 with a bell-shaped curve. The approximating curve is shown in Figure 11.22.

Distribution of Sample Proportions

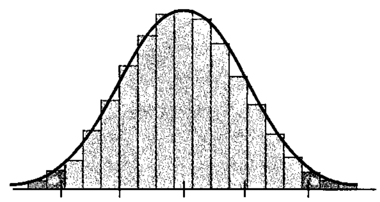


Figure 11.22

Figure 11.22 shows that the distribution of sample proportions can be modeled by a normal distribution. In fact, the bell-shaped curve in Figure 11.22 represents a normal distribution with $\mu = 0.3$ and $\sigma \approx 0.065$. Notice that the peaks of the histogram and the bell-shaped curve occur at $p = 0.3$. The location of these peaks tells us that if all possible samples of 50 students were polled, then the mean of the sample proportions would equal 30%, the population proportion.

We can model the distribution of sample proportions by a normal distribution if the sample size n is large enough. The mean of the approximating normal distribution equals the population proportion. It turns out that the standard deviation can be computed from the population proportion, p, and the sample size, n. We can also determine the value of n that is "large enough" to ensure that the distribution of sample proportions really is approximately normal. The next box describes what conditions must be met.

DISTRIBUTION OF SAMPLE PROPORTIONS

If samples of size n are taken from a population having a population proportion p, then the set of all sample proportions has mean and standard deviation given by

$$\text{mean} = p \quad \text{and} \quad \text{standard deviation} = \sqrt{\frac{p(1-p)}{n}}.$$

If *both* of the conditions

$$p - 3\sqrt{\frac{p(1-p)}{n}} > 0 \quad \text{and} \quad p + 3\sqrt{\frac{p(1-p)}{n}} < 1$$

are met, then n is large enough that the distribution of sample proportions, $\hat{p}$, can be treated as approximately normal.

EXAMPLE 11.15 Suppose that the population proportion of a group is 0.4, and we choose a simple random sample of size 30. Find the mean and standard deviation of the set of all the sample proportions.

SOLUTION In this example, $p = 0.4$ and $n = 30$. Substituting these values into the preceding formula, the set of all sample proportions has a mean of 0.4 and a standard deviation of

$$\sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{(0.4)(1-0.4)}{30}} = \sqrt{\frac{(0.4)(0.6)}{30}} = \sqrt{\frac{0.24}{30}} \approx 0.09.$$

Figure 11.23 shows a histogram for the sample proportions in this example.

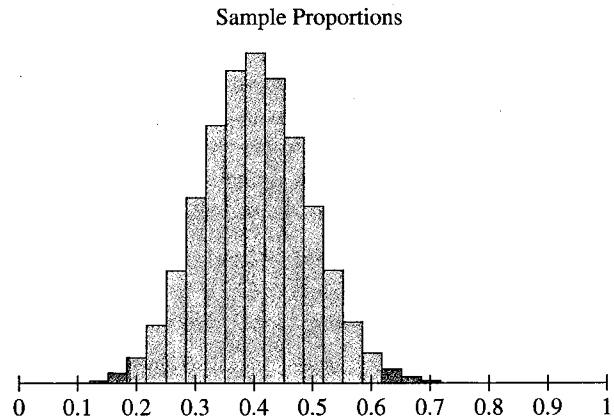


Figure 11.23

The sample size of $n = 30$ in Example 11.15 is considered "large" because

$$p - 3\sqrt{\frac{p(1-p)}{n}} = 0.40 - 3 \times 0.09 = 0.13 > 0$$

and

$$p + 3\sqrt{\frac{p(1-p)}{n}} = 0.40 + 3 \times 0.09 = 0.67 < 1$$

both hold. Thus, in this case a sample size of 30 is large enough that the distribution of sample proportions is approximately normal. With rare exceptions, as long as the sample contains a few dozen elements, we do not need to worry about whether n is large enough.

When we know that a histogram can be approximated by a bell-shaped curve, we can use what we have learned about finding areas under the bell-shaped standard normal curve to determine the percentage of data points that lie between any two values. In the earlier discussion of the President's budget proposal, the population was the set of registered voters in the United States, and the group of interest was the set of supporters of the President's budget. The sample size was $n = 413$, and the population proportion was $p = 0.5$. Therefore, the mean of the set of all sample proportions is 0.50, and the standard deviation is $\sqrt{\frac{(0.5)(1-0.5)}{413}} \approx 0.02460 \approx 0.025$.

In Section 11.2, we saw that a measurement in any normal distribution may be converted to a z-score that represents its position in the standard normal distribution. In the same way, a sample proportion may be converted to a z-score by subtracting the mean from the sample proportion and dividing the result by the standard deviation. For instance, in the case of the President's budget proposal, a sample proportion of 0.525 corresponds to a z-score of $z = \frac{0.525 - 0.5}{0.025} = 1$. The normal distribution of sample proportions is illustrated in Figure 11.24. It includes both a sample proportion scale and the z-score scale and shows a bell-shaped curve matched to the z-score scale.

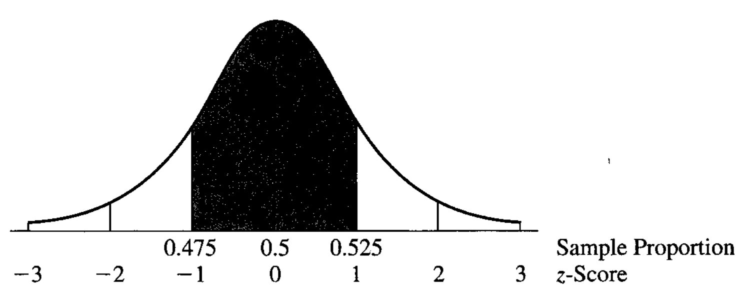


Figure 11.24

Figure 11.24 shows that the percentage of samples for which the sample proportion is between 47.5% and 52.5% is the same as the area between $z = -1$ and $z = 1$ under a standard normal curve. This area is 68%, or about $\frac{2}{3}$. We can conclude that of all possible samples of 413 registered U.S. voters, approximately $\frac{2}{3}$ of them will yield a sample proportion between 47.5% and 52.5% supporting the President's budget. Thus, it is not surprising that in this example we saw a sample proportion of 48%.

Sample proportions are frequently used in opinion surveys and political polls. The next example provides one illustration of a recent opinion poll conducted by Fox TV news.

EXAMPLE 11.16 In December 2002, according to www.pollingreport.com, Fox News sampled 900 registered voters nationwide and asked the question "If a smallpox vaccination were offered to you, would you take the shot or not?" Suppose it is known that 60% of all Americans would take the smallpox vaccination. Considering all possible polls of 900 voters, what is the approximate percentage of samples for which between 58% and 62% of voters in the sample would take the shot?

SOLUTION In this situation, the population consists of the set of U.S. registered voters, and the group of interest is the set of people who would choose to get a smallpox vaccination. The sample is the group of 900 voters surveyed by Fox News. Because we know that $p = 60\% = 0.6$, the set of all sample proportions has a mean of 0.6. The sample size is $n = 900$, so the standard deviation of the set of all possible sample proportions is

$$\sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{(0.6)(1-0.6)}{900}} = \sqrt{\frac{(0.6)(0.4)}{900}} \approx 0.02.$$

Thus, a sample proportion of $60\% - 2\% = 58\%$ has a z-score of -1, and $60\% + 2\% = 62\%$ has a z-score of $+1$. Figure 11.25 shows a bell-shaped curve that represents all possible sample proportions. It includes both a sample proportion scale and a z-score scale.

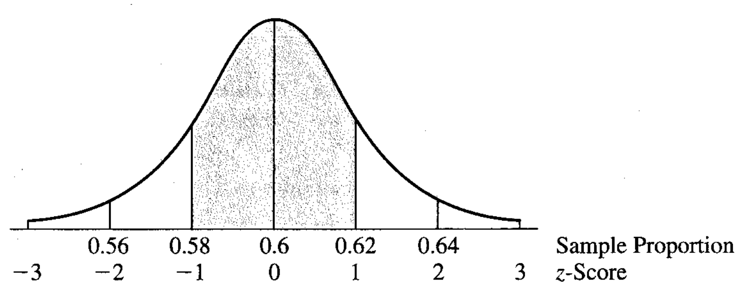


Figure 11.25

Figure 11.25 shows us that the percentage of samples between 0.58 and 0.62 is the same as the area under a standard normal curve between $z = -1$ and $z = 1$. By the 68–95–99.7 rule for normal distributions, this area is approximately 68%. In other words, of all possible polls of 900 American voters that could have been conducted by Fox News, in about 68% of those polls, between 58% and 62% of people polled would choose to be vaccinated against smallpox. In the actual 2002 Fox News poll, 59% of voters surveyed said "yes" to the question. ■

THE STANDARD ERROR

In the examples we have discussed thus far, the population proportion was given. However, if we already knew the population proportion, we would not need to take samples and compute the sample proportion to try to approximate the population proportion. In most situations, we have the opposite problem: we know only the sample proportion and want an estimate of how close it is to the true population proportion. When the poll was taken to determine the level of support for the President's budget, for example, the population proportion, p, was actually unknown. Indeed, researchers took the poll in order to determine the value of p. After they chose a sample, they computed the sample proportion, $\hat{p}$, to be 48% by dividing the number of people in the sample that supported the budget by 413, the sample size. Thus, 48% should be our best guess for the population proportion. The question is how good a guess is it?

The following general formula can help us answer this question.

> ***Definition***
>
> **STANDARD ERROR**
>
> If a representative sample of size n is taken from a population, and if the sample proportion equals $\hat{p}$, then the standard deviation of the set of all sample proportions is approximately
>
> $$\hat{s} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}},$$
>
> which is known as the **standard error** of the sample.

Notice that this formula for the standard error is the same as the standard deviation formula with p replaced by $\hat{p}$. The standard error is a very good approximation to the true standard deviation of the set of all sample proportions.

EXAMPLE 11.17 What is the standard error in a sample of size 400 if the sample proportion is 35%?

SOLUTION We use the preceding formula to compute the standard error.

$$\hat{s} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{(0.35)(1-0.35)}{400}} =$$

$$\sqrt{\frac{(0.35)(0.65)}{400}} = \sqrt{0.00056875} \approx 0.024. \blacksquare$$

CONFIDENCE INTERVALS AND THE MARGIN OF ERROR

Recall that for any normal distribution, 95% of the data must be within 2 standard deviations of the mean (by the 68–95–99.7 rule for normal distributions). Applying this characteristic of normal distributions to the distribution of sample proportions gives us some idea of how "good" an estimate of the true population proportion a sample proportion is. We can say that 95% of the time, or 19 out of 20 times, the sample proportion will be within 2 standard deviations of the population proportion. If the standard deviation is not known, we can use the standard error (which can be calculated) and conclude that in 95% of the cases the population proportion is within 2 standard errors of the sample proportion. We say that a **95% confidence interval** is the interval of numbers from $\hat{p} - 2\hat{s}$ to $\hat{p} + 2\hat{s}$ (Figure 11.26).

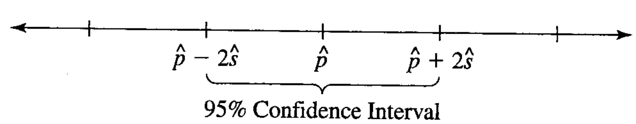


Figure 11.26

Any value in the interval from $\hat{p} - 2\hat{s}$ to $\hat{p} + 2\hat{s}$ is a reasonable estimate for the population proportion, p. This interval is called a 95% confidence interval because, for 95% of the samples, the interval computed in this way will contain p. Two reasonable locations of $\hat{p}$ are pictured in Figures 11.27(a) and (b).

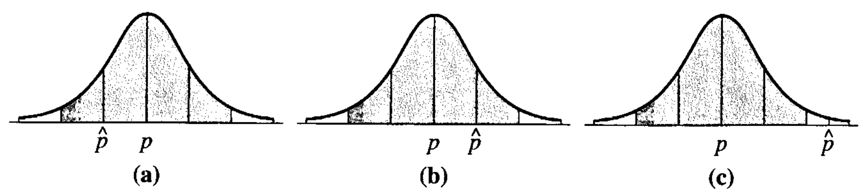


Figure 11.27

Any guess of a proportion p that is *not* in the 95% confidence interval is not a "good" guess [Figure 11.27(c)]. Such a guess is unlikely to be the population proportion because fewer than 5% of the samples taken would give an interval that did not contain p.

Returning to the example of the President's budget, the size of our sample was $n = 413$. The sample proportion was $\hat{p} = 0.48$. Thus, the standard error is

$$\hat{s} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{(0.48)(1-0.48)}{413}} =$$

$$\sqrt{\frac{(0.48)(0.52)}{413}} = \sqrt{\frac{0.2496}{413}} \approx 0.02458 \approx 0.025.$$

Recall that earlier we calculated a standard deviation of $0.02460 \approx 0.025$ for this situation using p in the formula rather than $\hat{p}$ as we do now. Notice that this new estimate is very close to the same value.

Our 95% confidence interval is from $\hat{p} - 2\hat{s} = 0.48 - 2(0.025) = 0.43$ to $\hat{p} + 2\hat{s} = 0.48 + 2(0.025) = 0.53$. This means that we have 95% confidence that the population proportion is between 0.43 and 0.53. In other words, of many different random samples of registered voters, 95% of them (19 of 20) will yield an interval that contains the population proportion.

The value $2\hat{s}$ is called the margin of error in the estimate of the population proportion. In the case of the poll of voter opinion on the President's budget, we found that $\hat{s} = 0.025$, so $2\hat{s} = 2(0.025) = 0.05$. Thus, we have confirmed that the poll has a margin of error of 5%. Usually the margin of error is described with the words "plus or minus," so for the poll of voter opinion on the President's budget, we say the margin of error was plus or minus 5% or, $\pm 5\%$.

Definition

95% CONFIDENCE INTERVAL AND MARGIN OF ERROR

If a sufficiently large representative sample has sample proportion $\hat{p}$ and standard error $\hat{s}$, then the **95% confidence interval** for the population proportion is the interval of numbers from $\hat{p} - 2\hat{s}$ to $\hat{p} + 2\hat{s}$. The **margin of error** for the confidence interval is $\pm 2\hat{s}$.

EXAMPLE 11.18 Determine a 95% confidence interval for a sample of size 400 with a sample proportion of 35%. What is the margin of error?

SOLUTION In Example 11.17, we computed the standard error of this sample to be $\hat{s} = 0.024 = 2.4\%$. The confidence interval therefore goes from 35% minus 2 standard errors to 35% plus 2 standard errors; that is, the 95% confidence interval is from

$$\hat{p} - 2\hat{s} = 35\% - 2(2.4\%) = 35\% - 4.8\% = 30.2\%$$

to

$$\hat{p} + 2\hat{s} = 35\% + 2(2.4\%) = 35\% + 4.8\% = 39.8\%.$$

The margin of error is plus or minus 2 standard errors, that is, $\pm 4.8\%$. ■

Note that if the values are rounded to full percentages, they are rounded "outward" so that at least 95% of the sample proportions are still included. In Example 11.18, we would state the 95% confidence interval as 30% to 40%. For a 95% confidence interval of 41.6% to 46.4%, the percentages would be rounded to 41% and 47%.

EXAMPLE 11.19 Suppose that we sample 600 U.S. citizens and ask them if they drive an American-built car as their primary source of transportation. In this sample of 600 people, 362 people say that they do. Compute a 95% confidence interval for the proportion of the population that drives an American-built car as their primary source of transportation. What is the margin of error?

SOLUTION The sample proportion is $\hat{p} = \frac{362}{600} = 0.603$. The standard error is computed as

$$\hat{s} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{(0.603)(1-0.603)}{600}} = \sqrt{\frac{(0.603)(0.397)}{600}} \approx 0.020.$$

A 95% confidence interval is the interval of numbers within 2 standard errors of $\hat{p}$. In this case, the confidence interval is the interval between $\hat{p} - 2\hat{s} = 0.603 - 2(0.020) = 0.563$ and $\hat{p} + 2\hat{s} = 0.603 + 2(0.020) = 0.643$ (Figure 11.28).

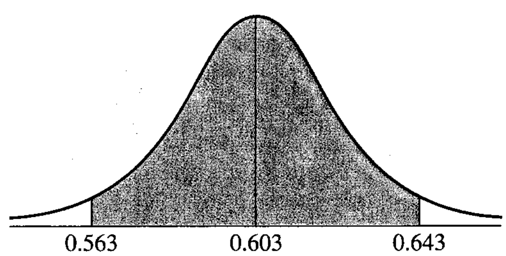


Figure 11.28

With a confidence level of 95%, we conclude that the true population proportion would lie between 56.3% and 64.3%. We could also report the results of this poll by stating that the percentage of U.S. citizens driving American-built cars as their primary transportation is 60.3%, with a margin of error of $\pm 4\%$. ■

Polling results are reported in the media in the same way as we interpreted the confidence interval obtained in Example 11.19. The following example gives one illustration.

EXAMPLE 11.20 According to www.pollingreport.com, an ABC News/*Washington Post* poll conducted on October 3, 2003, surveyed current attitudes about health care in the United States. On that day, 1000 adults nationwide were asked the question "Thinking about health care in the country as a whole, are you generally satisifed or dissatisfied with the quality of health care in this country?" Researchers reported that

44% of respondents said they were satisfied with the quality of health care in the United States. They also stated that the margin of error in the poll was ±3%. Assuming that the pollsters calculated a 95% confidence interval, which is common, verify that the stated margin of error is correct, and explain what it means.

SOLUTION For this particular poll we know that $n = 1000$ and $\hat{p} = 0.44$. To determine the margin of error, we must find the value of the standard error $\hat{s}$. Using the formula for the standard error, we find that

$$\hat{s} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{(0.44)(1-0.44)}{1000}} = \sqrt{\frac{(0.44)(0.56)}{1000}} \approx 0.0157.$$

The margin of error is $2\hat{s} \approx 2(0.0157) = 0.0314$, which is approximately 3%. So, the margin of error, as stated in the report, is correct.

By stating that the margin of error is 3%, the researchers are saying that they are 95% certain that the true percentage of U.S. adults who are satisfied with health care in the United States is between $44\% - 3\% = 41\%$ and $44\% + 3\% = 47\%$. In other words, a 95% confidence interval for the proportion of Americans who are satisfied with U.S. health care is the interval from 41% to 47%. ■

CONFIDENCE INTERVALS AND QUALITY CONTROL

The same reasoning we used in Example 11.19 also works with quality control problems. If we take a sample of items from a single day's production and check each one for defects, the proportion of defects in the sample gives the best estimate for the proportion of defects in the entire day's production. Testing some of the items from a production run is not much different from surveying a sample of people.

EXAMPLE 11.21

a. Suppose a manufacturer tests 1000 computers chips and finds that 216 of them are defective. Find a 95% confidence interval for the population proportion (or percentage) of defective chips.

b. Suppose a manufacturer tests 10,000 rather than 1000 computer chips and finds that 2160 of these are defective. What is the 95% confidence interval in this case? Does choosing a larger sample give significantly better results?

SOLUTION

a. In the first case, $n = 1000$ and $\hat{p} = \frac{216}{1000} = 0.216$. The standard error is then $\hat{s} = \sqrt{\frac{(0.216)(1-0.216)}{1000}} \approx 0.013$. Thus, the endpoints of our 95% confidence interval are $\hat{p} - 2\hat{s} = 0.216 - 2(0.013) = 0.190$ and $\hat{p} + 2\hat{s} = 0.216 + 2(0.013) = 0.242$. That is, we can say with 95% confidence that the true population proportion (percentage of defectives) lies between 19.0% and 24.2%.

b. Here, $n = 10{,}000$ and $\hat{p} = \frac{2160}{10{,}000} = 0.216$ as before. However, the standard error is now $\hat{s} = \sqrt{\frac{(0.216)(1-0.216)}{10{,}000}} \approx 0.004$. Thus, we have $\hat{p} - 2\hat{s} = 0.216 - 2(0.004) = 0.208$ and $\hat{p} + 2\hat{s} = 0.216 + 2(0.004) = 0.224$. A 95% confidence interval is the interval of proportions between 20.8% and 22.4%. Because the standard error is much smaller in this case, this confidence interval is narrower than the one calculated in part (a). And because the interval between 20.8% and 22.4% is narrower, it gives a significantly better estimate of the true percentage of defectives than when 1000 samples are taken. ■

SOLUTION OF THE INITIAL PROBLEM

A candy bar company has a promotion in which letters printed inside some of the wrappers earn the buyer a prize. Suppose that you buy 400 of these candy bars and you discover that 25 wrappers are winners. You are thinking of buying another 1000 candy bars. How many wrappers would you expect to have letters printed on the inside?

SOLUTION Consider the first 400 bars that you unwrapped. Here, $n = 400$. The sample proportion is $\hat{p} = \frac{25}{400} = 0.0625$. Thus, we might estimate that of the next 1000 wrappers, 6.25% of them, or $0.0625(1000) = 62.5$, might be winners.

On the other hand, we may use the standard error to create a confidence interval for the number of winners. The standard error here is $\hat{s} = \sqrt{\frac{0.0625(1 - 0.0625)}{400}} \approx 0.0121$. We calculate our 95% confidence interval as

$$\hat{p} - 2\hat{s} = 0.0625 - 2(0.0121) = 0.0383,$$

and

$$\hat{p} + 2\hat{s} = 0.0625 + 2(0.0121) = 0.0867.$$

A 95% confidence interval for this situation is the interval of numbers between $0.0383 = 3.83\%$ and $0.0867 = 8.67\%$. Thus, out of 1000 candy bars, you should expect between $1000(0.0383) \approx 38$ and $1000(0.0867) \approx 87$ to have letters printed on the inside.

PROBLEM SET 11.3

1. Use your calculator to find each of the following. Round to the nearest thousandth.

a. $\sqrt{\dfrac{(0.2)(0.8)}{50}}$

b. $\sqrt{\dfrac{(0.14)(1 - 0.14)}{634}}$

c. $0.49 - 3\sqrt{\dfrac{(0.49)(1 - 0.49)}{470}}$

2. Use a calculator to find each of the following. Round to the nearest thousandth.

a. $\sqrt{\dfrac{(0.3)(0.7)}{30}}$

b. $\sqrt{\dfrac{(0.81)(1 - 0.81)}{199}}$

c. $0.26 - 3\sqrt{\dfrac{0.26(1 - 0.26)}{38}}$

3. All possible samples of size 250 are taken from a population that has a population proportion of 0.58.

a. Find the mean of the set of sample proportions.

b. Find the standard deviation of the set of sample proportions.

4. All possible samples of size 1358 are taken from a population that has a population proportion of 0.61.

a. Find the mean of the set of sample proportions.

b. Find the standard deviation of the set of sample proportions.

5. A factory produces 6000 cars during a certain week, and 300 of them have significant problems needing correction. Inspectors select 60 of the cars for a detailed inspection and 5 have a problem needing correction.

a. What is the population and what is the sample?

b. What is the population proportion of cars having problems?

c. What is the sample proportion of cars having problems?

6. A manufacturer creates an assortment of candies by mixing 500 caramels with 1000 chocolate-covered nuts. These are then put into half-pound packages. A package is opened and found to have 12 caramels and 18 chocolate-covered nuts.

a. What is the population and what is the sample?

b. What is the population proportion of caramels?

c. What is the sample proportion of caramels?

7. Suppose a certain city has 7140 registered voters, of which 3460 are Democrats, 3250 are Republicans, and 430 are Independents. A pre-election canvassing in a given neighborhood reveals the following numbers of registered voters: 185 Democrats, 210 Republicans, and 25 Independents.

a. What is the population and what is the sample?

b. Find the sample size.

c. What is the population proportion of registered Republicans?

d. What is the sample proportion of registered Republicans?

8. According to the 2000 U.S. Census, the number of Nevada residents age 25 or older was 1,310,176. Of those, the number who had earned at most a bachelor's degree was 158,078. Suppose a research group conducted a telephone survey of Nevada residents age 25 or older and it was determined that 299 had earned at most an associate's degree, 45 had earned at most a bachelor's degree, and 22 had earned a graduate degree.

a. What is the population and what is the sample?

b. Find the sample size.

c. What is the population proportion of people who earned at most a bachelor's degree?

d. What is the sample proportion of people who earned at most a bachelor's degree?

9. A factory produces 6000 boxes of cereal during a certain day and significantly underfills 300 of them. An inspector randomly selects a sample of 55 boxes for a detailed weighing.

a. Find the population proportion of boxes of cereal that are significantly underfilled.

b. Find the mean and standard deviation of the set of all sample proportions of underfilled boxes.

c. Is the sample size large enough so that we can conclude the distribution of sample proportions is approximately normal? Why or why not?

10. Out of the 3,450,000 registered voters in a state, 98,549 favor a moratorium on state executions until death penalty procedures are officially reviewed. A reporter will take a sample of 15 registered voters and ask each of them whether a moratorium is favored.

a. Find the population proportion of those who favor the moratorium.

b. Find the mean and standard deviation of the set of all the sample proportions of those who favor a moratorium.

c. Is the sample size large enough so that we can conclude the distribution of sample proportions is approximately normal? Why or why not?

11. Suppose that out of 27,560 people who live in a certain city, 22,048 support the continuation of the manned space shuttle program.

a. Find the population proportion of people who support the continuation of the manned space shuttle program.

b. Find the mean and standard deviation of the set of all the sample proportions when samples of size 150 are taken from the population.

c. Find the mean and standard deviation of the set of all the sample proportions when samples of size 3000 are taken from the population.

d. Compare the results from parts (b) and (c) and explain what you observe.

12. On July 1, 2002, the population of Oregon was estimated to be 3,504,700. Suppose that 2,278,055 Oregonians would not support a tax increase to fund schools.

a. Find the population proportion of Oregonians who would not support a tax increase to fund schools.

b. Find the mean and standard deviation of the set of all the sample proportions when samples of size 2400 are taken from the population.

c. Find the mean and standard deviation of the set of all the sample proportions when samples of size 10,000 are taken from the population.

d. Compare the results from parts (b) and (c) and explain what you observe.

13. Explain the difference between p and $\hat{p}$.

14. Explain the difference between the standard deviation of the set of all sample proportions and the standard error of the sample.

15. Suppose a student has five courses to complete as requirements in science and humanities and decides to take three of them next term. Since he registers early and there are multiple sections for each course, he feels free to choose any three of the five. The required courses are Science A (SA), Science B (SB), Humanities A (HA), Humanities B (HB), and Humanities C (HC). There are 10 ways in which he can select three of the five courses. Consider each selection of three courses as one sample.

a. Find the population proportion of required humanities courses he can take.

b. The following table contains his possible course selections. Find the sample proportion of required humanities courses for each selection (sample) of three classes.

Course Selections	Sample Proportion of Humanities Courses
SA, SB, HA	
SA, SB, HB	
SA, SB, HC	
SA, HA, HB	
SA, HA, HC	
SA, HB, HC	
SB, HA, HB	
SB, HA, HC	
SB, HB, HC	
HA, HB, HC	

c. Find the mean of the sample proportions and compare it to the population proportion. What do you observe?

16. Six students tie for top honors in a graduating class. The administration has decided that the top four students will give speeches during the graduation ceremonies, so school officials decide to pick four of the students at random. The six students are Ann (A), Betty (B), Carol (C), Diana (D), Eddie (E), and Fred (F). There are 15 ways in which the administration can select four students out of the six.

a. Find the population proportion of females among the top 6 students.

b. The following table contains the possible student selections. Find the sample proportion of females for each selection (sample) of four students.

Students	Sample Proportion of Females
A, B, C, D	
A, B, C, E	
A, B, C, F	
A, B, D, E	
A, B, D, F	
A, B, E, F	
A, C, D, E	
A, C, D, F	
A, C, E, F	
A, D, E, F	
B, C, D, E	
B, C, E, F	
B, C, D, F	
B, D, E, F	
C, D, E, F	

c. Find the mean of the sample proportions and compare it to the population proportion. What do you observe?

17. Find the standard error in a sample of size 40, given each of the following sample proportions.

a. 10% **b.** 25% **c.** 50% **d.** 80%

18. Find the standard error in a sample of size 300, given each of the following sample proportions.

a. 10% **b.** 25% **c.** 50% **d.** 80%

19. Find the standard error for a sample proportion of 45%, given each of the following sample sizes.

a. 50 **b.** 100 **c.** 500 **d.** 1000

20. Find the standard error for a sample proportion of 90%, given each of the following sample sizes.

a. 50 **b.** 100 **c.** 500 **d.** 1000

21. Five hundred students at a high school are randomly selected for a student services survey. Of those selected, 265 are females. Find the standard error for the proportion of females. Interpret the meaning of the standard error in this case.

22. Sixty cars are randomly selected from the weekly production of cars at a plant. Five of these cars have problems requiring corrections. Find the standard error for the proportion of cars requiring corrections. Interpret the meaning of the standard error in this case.

23. In a September 2003 Virginia Commonwealth University Life Sciences Survey, adults nationwide were asked, "In general, do you think that it is morally acceptable or morally wrong to use human cloning technology in developing new treatments for disease?" Of the 1003 adults who were contacted, 361 thought it was morally acceptable. Find the sample proportion and the standard error for the proportion of adults who thought it was morally acceptable. Explain how you would interpret these two values. (*Source:* www.pollingreport.com.)

24. In a January 2003 CNN/*USA Today*/Gallup Poll, adults nationwide were asked the question, "Do you think that cloning that is designed specifically to result in the birth of a human being should be legal or illegal in the United States?" Of the 1000 adults surveyed, 110 said that it should be legal. Find the sample proportion and the standard error for the proportion of adults who thought it should be legal. Explain how you would interpret these two values. (*Source:* www.pollingreport.com.)

25. A general biology class is doing a project on physical characteristics. The students in the class were randomly assigned to the section after preregistration, so the instructor considers them to be a random sample of the student body. There are 42 students in the class, and 7 are left-handed.

a. Find the sample proportion and standard error for the proportion of students who are left-handed.

b. Find a 95% confidence interval for the population proportion of left-handed students.

c. The instructor made an assumption that the students in the class were representative of the student body. Explain whether that assumption is reasonable.

26. An instructor for a government studies course at a community college surveys the 35 students in her class, asking whether they would approve of a tax increase to fund social programs. The students in the class range in age from 19 to 57 and seem to come from different social classes, so the instructor considers them to be a representative sample of adults in the city. Of the 35 students surveyed, 9 approved of a tax increase to fund social programs.

a. Find the sample proportion and standard error for the proportion of adults who approve of a tax increase to fund social programs.

b. Find a 95% confidence interval for the population proportion of adults who approve of a tax increase to fund social programs.

c. The instructor made an assumption that the students in the class were representative of adults in the city. Explain whether that assumption is reasonable.

27. If a news organization reported that 39% of Americans believe the speed limit on all freeways should be raised to 85 miles per hour and gave the margin of error as 4%, what is the standard error of the sample? How big was the sample?

28. If a news organization reported that 47% of Americans believe the Alaskan Arctic Wildlife Refuge should be opened to oil and gas exploration and gave the margin of error as 3%, what is the standard error of the sample? How big was the sample?

29. The student services office of a university is concerned about student acceptance of new registration procedures. A random sample of students is selected and contacted. They are asked whether they find the new procedures satisfactory. Of the 280 students who respond, 172 are satisfied with the new procedures.

a. Find the sample proportion and standard error.

b. Find the 95% confidence interval for the percentage of students who are satisfied with the new registration procedures.

c. Explain how to interpret the 95% confidence interval.

30. A company that produces flashlight batteries wishes to know what percentage of its batteries will last longer than 30 hours. It selects and tests a random sample of 1000 batteries. Of these batteries, 917 last 30 hours or more.

a. Find the sample proportion and standard error.

b. Find the 95% confidence interval for the percentage of batteries that last 30 hours or more.

c. Explain how to interpret the 95% confidence interval.

31. In a February 2003 *Los Angeles Times* opinion poll, out of a sample of 1385 adults nationwide, 54% approved of President Bush's proposal to allow individuals to divert part of their Social Security payroll taxes into private accounts, which they could personally invest in the stock market for their retirement. (*Source:* www.pollingreport.com.)

a. Determine a 95% confidence interval for the population proportion of adults who favor President Bush's proposal.

b. Find the margin of error and explain its meaning.

32. In a February 2003 *Los Angeles Times* opinion poll, out of a sample of 1385 adults nationwide, 43% supported a complete ban on all research into human cloning, without exception. (*Source:* www.pollingreport.com.)

a. Determine a 95% confidence interval for the proportion of adults who support a complete ban on all research into human cloning.

b. Find the margin of error and explain its meaning.

33. Administrators at a college are interested in the number of students who are working 10 or more hours per week while taking full-time class loads. A random sample of 240 full-time students reveals that 105 of the students are working 10 or more hours per week. Find a 95% confidence interval for the percentage of full-time students who are working 10 or more hours per week.

34. A random survey of 500 pregnant women conducted in a large northeastern city indicated that 145 of them preferred a female obstetrician to a male obstetrician. Find a 95% confidence interval for the percentage of pregnant women in the city who would prefer a female obstetrician.

35. Title IX is a federal law that prohibits discrimination on the basis of gender in any high school or college that receives federal funds. It has been used to ensure that women have equal opportunities in high school and college athletics and are not discriminated against. In a January 2003 *Wall Street Journal* poll, out of 500 adults surveyed nationwide, 68% approve of Title IX. Find the 95% confidence interval for the proportion of people who approve of Title IX. What is the margin of error?

36. In an effort to comply with the requirements of Title IX (problem 35), many schools have had to cut funding for men's athletics in order to increase funding for women's programs. In a January 2003 *Wall Street Journal* poll, out of 500 adults surveyed nationwide, 27% disapprove of cutbacks in men's athletics. Find the 95% confidence interval for the population proportion of people who disapprove of cutbacks in men's athletics for the purpose of satisfying the requirements of Title IX. What is the margin of error?

37. The United States has not had a military draft since 1970, and it is suspected that only about 15% of the population would favor its reinstatement. In order to estimate the proportion of people in the United States who would favor reinstating the draft, we must take a sample.

a. How large a sample must be taken to have a 5% margin of error?

b. How large a sample must be taken to have a 1% margin of error?

38. A woman has never held the position of President of the United States, and it is believed that 75% of Americans would be willing to vote for a female candidate for president if she were qualified for the job. In order to estimate the proportion of Americans who would be willing to vote for a female candidate for president if she were qualified for the job, we must take a sample.

a. How large a sample must be taken to have a 5% margin of error?

b. How large a sample must be taken to have a 1% margin of error?

Problems 39 through 42

Public opinion polls typically report a margin of error of 5% or less. The margin of error gives a value to the uncertainty about survey results. It is not surprising, then, that one important factor in determining the margin of error is the sample size.

39. Consider a sample proportion of 0.5. Notice in this exercise how the sample sizes and confidence intervals change as the margin of error decreases:

a. Fill in the following table.

Margin of of Error	$\hat{s}$	Sample Size $n = \dfrac{\hat{p}(1-\hat{p})}{\hat{s}^2}$	95% Confidence Interval
10%			
5%			
1%			
0.1%			

b. Compare the sample size requirements and confidence intervals for the various margins of error in part (a). What do you conclude?

40. Consider a sample proportion of 0.4. Notice in this exercise how the sample sizes and confidence intervals change as the margin of error decreases:

a. Fill in the following table.

Margin of of Error	$\hat{s}$	Sample Size $n = \frac{\hat{p}(1-\hat{p})}{\hat{s}^2}$	95% Confidence Interval
10%			
5%			
1%			
0.1%			

b. Compare the sample size requirements and confidence intervals for the various margins of error in part (a). What do you conclude?

41. Notice in this exercise how the margin of error changes as the sample proportion, $\hat{p}$, and the sample size, n, change.

a. For each combination of sample proportion and sample size in the following table, find the margin of error.

Margin of Error for Given Values of n and $\hat{p}$.				
	$n = 30$	$n = 50$	$n = 100$	$n = 500$
$\hat{p} = 0.25$				
$\hat{p} = 0.3$				
$\hat{p} = 0.5$				

b. Consider the margins of error from part (a). Discuss whether n or $\hat{p}$ causes the greatest change in the margin of error.

42. Notice in this exercise how the margin of error changes as the sample proportion, $\hat{p}$, and the sample size, n, change.

a. For each combination of sample proportion and sample size in the following table, find the margin of error.

Margin of Error for Given Values of n and $\hat{p}$.				
	$n = 100$	$n = 500$	$n = 1000$	$n = 3000$
$\hat{p} = 0.2$				
$\hat{p} = 0.6$				
$\hat{p} = 0.9$				

b. Consider the margins of error from part (a). Discuss whether n or $\hat{p}$ causes the greatest change in the margin of error.

Extended Problems

43. Nielsen Media Research, a television ratings company, randomly selects 5000 "Nielsen Families" and carefully monitors their television viewing. Television shows are renewed or canceled based on Nielsen Family viewing habits. For sample sizes of 5000, find the margin of error for each of the following sample proportion values: 0.1, 0.3, 0.5, 0.7, and 0.9. Explain how only 5000 Nielsen Families can accurately reflect the television viewing habits of the entire population of the United States. (*Source:* www.nielsenmedia.com.)

44. The Gallup Organization and other polling companies routinely use sample sizes of 1000 to 1500. Polling companies must balance accuracy with the cost of increasing the sample size. For a sample proportion of 0.4, find the margins of error that result from sample sizes of 1000, 2000, 4000, and 8000. Suppose that doubling the sample size also doubles the cost to conduct the poll. Comment on the increased cost versus the increased accuracy that results from taking a larger sample.

Problems 45 and 46

In previous problems, we have been concerned with finding a 95% confidence interval for a population proportion. Although a 95% confidence interval is most common (public opinion polls use it almost exclusively), other confidence intervals are easy to define and calculate.

We know that a 95% confidence interval contains the values that are within 2 standard errors of the sample proportion because the distribution of sample proportions is assumed to be normal, and 95% of the sample proportions are within 2 standard errors (standard deviations) of the population proportion. Similarly, we can define a 99.7% confidence interval, which would be based on 3 standard errors, since 99.7% of all sample proportions are within 3 standard errors of the population proportion. Other commonly used confidence intervals are a 99% confidence interval based on 2.58 standard errors and a 90% confidence interval based on 1.65 standard errors.

45. Refer to Problem 31.

- **a.** Find a 90% confidence interval for the proportion of adults who favor President Bush's proposal.
- **b.** Find a 99.7% confidence interval for the proportion of adults who favor President Bush's proposal.
- **c.** Interpret and compare the intervals from parts (a) and (b).

46. Refer to Problem 32.

- **a.** Find a 90% confidence interval for the proportion of adults who support a complete ban on all research into human cloning.
- **b.** Find a 99% confidence interval for the proportion of adults who support a complete ban on all research into human cloning.
- **c.** Interpret and compare the intervals from parts (a) and (b).

47. By the time George Gallup started his American Institute of Public Opinion in 1935, opinion polling had been around for over 100 years. Research the history of public opinion polls. When and for what purpose were opinion polls conducted originally? What is a "straw poll"? How have opinion polls changed over the years to become more scientific? For more information, on the Internet search keywords "history of opinion polls." Write a report to summarize your findings.

48. Opinion polls often survey a random sample of people nationwide and consider the opinions of the people in the sample as representative of the entire population. If the sample size is large enough, the distribution of the set of all the sample proportions is approximately normal and we can calculate percentages and confidence intervals using the normal distribution. Let's see how these concepts work in practice. The math class you are attending will be the population. Survey every person who is registered for the class and ask their opinion about a topic of interest to you. Some ideas for the question include, "Do you support a ban on cell phone use while driving?", "Do you drink coffee?", or "Do you have Internet access at home?" Create a paper ballot, such as the following, for each student.

OPINION POLL

Do you support a ban on cell phone use while driving?

Yes ________ No ________

- **a.** Find the proportion of students in your class who favor a ban on cell phone use while driving. This value is the true population proportion.
- **b.** Using the results from part (a), find the mean and standard deviation of the set of all sample proportions of students who favor a ban on cell phone use while driving.
- **c.** Find the smallest sample size required so that the distribution of the set of sample proportions can be assumed to be approximately normal. Recall that the sample size will be considered sufficiently large if *both* of the following conditions are met

$$p - 3\sqrt{\frac{p(1-p)}{n}} > 0$$

and

$$p + 3\sqrt{\frac{p(1-p)}{n}} < 1.$$

- **d.** Thoroughly mix the ballots from the students in your class and randomly select a sample using the size you determined in part (c). Find the sample proportion of students who favor a ban on cell phone use while driving. Find a 95% confidence interval and calculate the margin of error. Interpret your results.

e. Repeat part (d) 10 times. What percent of confidence intervals did actually contain the true population proportion of students who favored a ban on cell phone use while driving?

f. Explain whether the results of your opinion poll can be used to represent the population of students in your school, your city, your state, or the nation.

49. A company decides to offer a "double your money back" guarantee on its high intensity beam flashlight. The flashlight costs \$15, and the company promises to refund \$30 to any customer who purchases a defective flashlight. To determine how much they might expect to pay out on this guarantee, the company tests 800 flashlights from a random sample. Of these, 28 are defective, and the rest are of high quality. Find a 99% confidence interval for the proportion of defective flashlights that are manufactured. Use this result to determine high and low estimates for the amount the company can reasonably expect to have to pay because of their stated guarantee. Suppose 10,000 flashlights will be sold with this guarantee.

50. Many websites for news organizations and polling organizations give information about survey results, sample sizes and margins of error. One such website is www.pollingreport.com. Using the survey results for five different surveys, construct the 95% confidence intervals and explain how each confidence interval should be interpreted with respect to each survey's question.

CHAPTER 11 REVIEW

Key Ideas and Questions

The following questions review the main ideas of this chapter. Write your answers to the questions and then refer to the pages listed by number to make certain that you have mastered these ideas.

1. What are the characteristics of a normal distribution? pgs. 703–704 What determines the shape of a normal distribution? pg. 704 What are the mean and standard deviation for a standard normal distribution? pg. 705
2. How can the symmetry of the standard normal curve be used to find areas under the curve? pg. 708 How can a table be used to calculate areas under a standard normal curve between two values? pgs. 707–709
3. What is the relationship between any normal distribution and the standard normal distribution? pg. 716 What does the 68–95–99.7 rule say about normal distributions? pg. 717
4. How is the population z-score calculated? pg. 719 How is the population z-score interpreted? pgs. 719–720 How do we use the tables for the standard normal distribution to calculate areas under *any* normal distribution? pgs. 719–720
5. What is the difference between the population proportion and the sample proportion? pg. 729 For a large enough sample size, what is true about the distribution of sample proportions? pg. 731 Why does sampling allow for meaningful results to come from surveys of a relatively small number of people? pgs. 731–732
6. How is the standard error calculated? pg. 734 What can be approximated by the standard error? pg. 734
7. How can a 95% confidence interval be interpreted? pg. 735 What is the margin of error? pg. 735
8. How does increasing the sample size affect the 95% confidence interval? pg. 737

Vocabulary

Following is a list of key vocabulary for this chapter. Mentally review each of these terms, write down the meaning of each one in your own words, and use it in a sentence. Then refer to the page number following each term to review any material that you are unsure of before solving the Chapter 11 Review Problems.

SECTION 11.1

Statistical Inference 701
Normal Distribution/Normally Distributed 703
Standard Normal Distribution 705
Tables for the Standard Normal Distribution 707–709

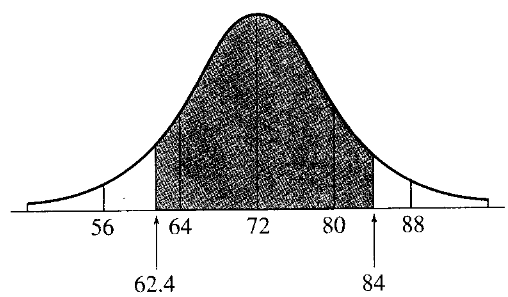
SECTION 11.2

68–95–99.7 Rule for Normal Distributions 717
Population z-score 719

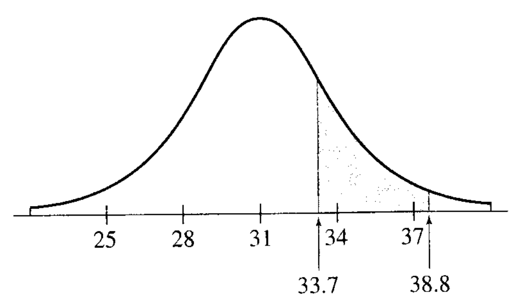
SECTION 11.3

Population Proportion 729
Sample Proportion 729
Distribution of Sample Proportions 731
Standard Error 734
95% Confidence Interval 735
Margin of Error 735

1. Consider a standard normal distribution.
   a. What percentage of values are within 1 standard deviation of the mean?
   b. Find the percentage of values that are between -2 and 2.
2. Consider a normal distribution with mean equal to μ and standard deviation equal to σ. Use the 68–95–99.7 rule for normal distributions.
   a. Find the percentage of values that are between $\mu - 2\sigma$ and $\mu + 2\sigma$.
   b. Find the percentage of values that are greater than μ.
   c. Find the percentage of values that are greater than $\mu - \sigma$ but less than $\mu + 2\sigma$.
3. a. What percentage of a standard normal population has values between -1 and 3?
   b. What percentage of a standard normal population has values greater than 3?
   c. What percentage of a standard normal population has values less than -1?
   d. Why do the percentages from parts (a), (b), and (c) add to 100%?
4. a. What percentage of a standard normal population has a value between -1.2 and 2.7?
   b. What percentage of a standard normal population has a value between 0.7 and 2.1?
   c. What percentage of a standard normal population has a value greater than 1.6?
5. Suppose that measurements made on a sample from a population are 16.1, 21.9, 22.3, and 18.6. Find the population z-scores for each of these measurements if the measurements are normally distributed with a mean of 17.9 and a standard deviation of 1.4. Round each z-score to the nearest tenth.
6. Suppose foot lengths of adult men are normally distributed with a mean of 11 inches and a standard deviation of 0.75 inches. Suppose foot lengths of adult women are normally distributed with a mean of 9 inches and a standard deviation of 1.2 inches.
   a. Sketch the normal curve that represents the foot lengths of adult men and shade the region from 9.5 inches to 12.125 inches. What percentages of measurements are between 9.5 and 12.125?
   b. Sketch the normal curve that represents the foot lengths of adult women and shade the region from 6 inches to 9.6 inches. What percentages of measurements are between 6 and 9.6?
   c. Compare the two normal curves from parts (a) and (b). Which of the two curves is taller and thinner? Which of the two curves is centered to the right of the other? Explain.
7. If measurements are taken from a normal distribution with a mean of 517, and approximately 99.74% of measurements are between 421 and 613, find the standard deviation.
8. Find the area of the shaded region for each of the following normal distributions.
   a. The mean is 72, and the standard deviation is 8.



   b. The mean is 31, and the standard deviation is 3.



9. Suppose that the total number of exams taken by students in college has a normal distribution with a mean of 54.5 and a standard deviation of 8.4.

a. What percentage of students take fewer than 40 exams?

b. What percentage of students take between 50 and 60 exams?

c. What percentage of students take more than 60 exams?

10. The Dolch list consists of the 220 words most frequently found among words read by children. Children usually learn these words in first or second grade. Most of the words cannot be read using basic decoding rules so they must be memorized as sight words. Assume that the number of Dolch sight words per minute that first-grade children can read is approximately normally distributed with a mean of 41 words per minute and a standard deviation of 9.19 words per minute. Round z-scores to the nearest tenth. (*Source:* www.whitworth.edu/Academic/Department/Education/MIT/Research/PDFDocuments/Journal/StephanieFlaherty.pdf.)

a. What percentage of first-grade children can read fewer than 25 sight words per minute?

b. What percentage of first-grade children can read more than 50 sight words per minute?

c. If a first-grade child is selected at random, what is the probability that the child will be able to read between 45 and 65 sight words per minute?

d. If a first-grade child is selected at random, what is the probability that child will be able to read between 10 and 35 sight words per minute?

Problems 11 and 12

The ACT is a national college admission examination. It is designed to assess a student's general education development and ability to complete college-level work. The ACT is made up of four tests in the following areas: mathematics, reading, science, and English. Scores for the ACT are approximately normally distributed. The following table contains test results for the 2003 ACT, which was taken by 1,175,059 students. (*Source:* www.act.org/aap/.)

2003 Results	Mathematics	Reading	Science	English	Composite
Mean	20.6	21.2	20.8	20.3	20.8
Standard Deviation	5.1	6.1	4.6	5.8	4.8

11. Colleges and Universities use the ACT as an admissions guide.

a. The University of Virginia's College at Wise gives preference to students who, among other things, have an ACT composite score of at least 18. What percentage of students from the 2003 group would be preferred by the University of Virginia's College at Wise? (*Source:* www.wise.virginia.edu/.)

b. The University of Missouri–Columbia requires an ACT composite score of at least 24. Approximately how many of the students who took the 2003 ACT could apply? (*Source:* http://admissions.missouri.edu/.)

c. For each of the four sections of the ACT (mathematics, reading, science, and english), what percentage of students scored between 20 and 24? What conclusion can you draw based on those percentages?

12. If a student scores at least a 24 on the science portion of the ACT, he or she is generally considered to have displayed a readiness for college biology. Similarly, if a student scores at least a 22 on the mathematics portion of the ACT, he or she is thought to be ready for college algebra.

a. What percentage of the students who took the 2003 ACT scored at least 24 on the science portion?

b. What percentage of the students who took the 2003 ACT scored at least 22 on the mathematics portion?

c. Compare parts (a) and (b). What do you conclude?

13. You take a sample of 30 candies from a jar of candies that have been well mixed. Suppose that 9 of these turn out to be chocolate. If there are 200 candies in the jar, what is the best estimate for the total number of chocolate candies in the jar?

14. A tub contains 20,000 jelly beans, 5000 of which are cherry-flavored. Suppose that a random sample of 48 jelly beans is taken and 9 are found to be cherry-flavored.

a. Identify the population and the sample.

b. What is the population proportion of cherry jelly beans?

c. What are the mean and standard deviation of the set of all the sample proportions of cherry jelly beans for samples of size 48?

d. Is the sample large enough that the distribution of sample proportions can be assumed to be normal? Justify your response.

e. Use the results of the sample to construct a 95% confidence interval for the proportion of cherry-flavored jelly beans in the jar.

15. Consider the jar of jelly beans from problem 14. Suppose a random sample of size 15 is taken from the jar and 1 jelly bean in the sample is cherry-flavored. Is the sample large enough that the distribution of sample proportions can be assumed to be normally distributed?

16. Suppose a school task force took a random sample of 300 families with school-age children in Massachusetts and found that 34 families send their children to private school.

a. Find the sample proportion of families with school-age children who send their children to private school.

b. Find and interpret the standard error.

c. Find a 95% confidence interval for the proportion of families with school-age children who send their children to private school.

17. In 2003, suppose 11% of computer users had Internet service through a cable modem. Suppose random samples of size 100 are taken.

a. Find the mean and standard deviation for the set of all the sample proportions.

b. What percent of the samples would have a sample proportion less than 8%?

c. What percent of the samples would have a sample proportion greater than 15%?

d. Suppose 21 computer users in the random sample of 100 had Internet service through a cable modem. Comment on the likelihood of obtaining a sample such that at least 21 computer users have cable Internet service.

e. Suppose an Internet provider took a random sample of 100 computer users and found that 8 have Internet service through a cable modem. Construct a 95% confidence interval for the proportion of computer users who have Internet service through a cable modem.

f. Suppose the Internet provider took a different random sample of 100 computer users and found that 10 have Internet service through a cable modem. Construct a 95% confidence interval for the proportion of computer users who have Internet service through a cable modem.

g. Compare the confidence intervals obtained in parts (e) and (f) to the true proportion. Comment on what you observe.

18. The formula for finding the standard deviation for the distribution of the set of all the sample proportions based on the sample size, n, is $s = \sqrt{\dfrac{p(1-p)}{n}}$.

a. Show that $n = \dfrac{p(1-p)}{s^2}$.

b. If we assume that $p = 0.5$, what minimum sample size is needed to obtain a value of $s = 0.01$?

c. If we assume that $p = 0.2$, what minimum sample size is needed to obtain a value of $s = 0.01$?

19. Suppose a representative sample of 100 people is surveyed and 62% feel that they pay too much money in taxes. Suppose the polling company samples the same population again, but this time they survey 1000 people, and they find that 60% feel that they pay too much money in taxes.

a. Find and compare the margins of error in each case.

b. Find and compare the 95% confidence intervals in each case.

c. Suppose the population proportion is 0.61. How large a sample would have to be taken to have a margin of error of 1%? Discuss the gain in accuracy in comparison to the required increase in sample size.

20. Each month, the U.S. Department of Labor conducts a survey to determine (among other things) the U.S. unemployment rate. In September 2003, the department reported that of the 146,545 civilians it surveyed, 8973 were unemployed. (*Source:* www.dol.gov.)

a. Construct and interpret a 95% confidence interval for the "true" unemployment rate.

b. Would a sample size of 100 have been large enough to assume the distribution of all sample proportions was normal? Explain.

21. In an October 2003 CNN/*USA Today*/Gallup Poll, Americans age 18 and older were asked the question, "Do you approve or disapprove of the way George W. Bush is handling his job as president?" Out of the 1006 Americans surveyed, 533 said they approved of the way George W. Bush was handling his job as president.

a. Identify the population and the sample.

b. Find and interpret the margin of error for the poll.

c. Construct a 95% confidence interval for the percentage of Americans who approved of the way George W. Bush was handling his job as president.

22. Suppose an Eastern university surveyed its students about their use of credit cards. Of the 840 students who responded, 605 said they had at least two credit cards, and 218 said they had missed payments.

a. Construct a 95% confidence interval for the percentage of students who have at least two credit cards.

b. Construct a 95% confidence interval for the percentage of students who have missed payments.

23. An opinion polling company conducted three random samples of individuals from a population and recorded the number of positive responses. Consider the following survey results:

Survey 1:	Number polled	200
	Positive responses	120
Survey 2:	Number polled	800
	Positive responses	480
Survey 3:	Number polled	1800
	Positive responses	1080

a. Find the margins of error for surveys 1, 2, and 3.

b. Compare the number polled in survey 2 to the number polled in survey 1, and compare the margin of error in survey 2 to the margin of error in survey 1. What do you notice?

c. Compare the number polled in survey 3 to the number polled in survey 1, and compare the margin of error in survey 3 to the margin of error in survey 1. What do you notice?

24. A researcher believes that males and females will do equally well on a test she has prepared. She administers the test to 360 males and 400 females. Out of those tested, 224 of the males and 282 of the females passed. The researcher said that although the percentage of females passing the test (70.5%) was higher than the percentage of males passing the test (62.2%), there was no conclusive evidence that females, in general, do better than males on the test.

a. Construct 95% confidence intervals for the percentage of females and the percentage of males passing the test.

b. Provide an argument supporting the researcher's conclusion.