CHAPTER 11

Inferential Statistics

Then

In Chapter 0, you found measures of center and spread and organized statistical data.

Now

In Chapter 11, you will:

- Use the shape of a distribution to select appropriate descriptive statistics.
- Construct and use probability distributions.
- Use the Central Limit Theorem.
- Find and use confidence intervals, and perform hypothesis testing.
- Analyze and predict using bivariate data.

Why?

ENVIRONMENTAL ENGINEERING Statistics are extremely important in engineering. In environmental engineering, hypothesis testing can be used to determine if a change in an emission level for a chemical has a significant impact on overall pollution. Also, confidence intervals can be used to help suggest restrictions on by-product wastes in ground water.

PREREAD Scan the study guide and review and use it to make two or three predictions about what you will learn in Chapter 11.

Noel Hendrickson/Getty images

connectED.mcgraw-hill.com Your Digital Math Portal

Animation | Vocabulary | eGlossary | Personal Tutor | Graphing Calculator | Audio | Self-Check Practice | Worksheets

Get Ready for the Chapter

Diagnose Readiness You have two options for checking Prerequisite Skills.

1 Textbook Option Take the Quick Check below.

QuickCheck

Find each value. (Lesson 0-7)

1. ${}_5P_2$ **2.** ${}_9P_4$ **3.** ${}_8C_3$

4. INTERNET The table shows the survey results of 18 high school students who were asked how many hours they spent on the Internet the previous week. (Lesson 0-8)

Hours Spent on the Internet					
2	3.5	1	8	2.5	7.5
10	4	5.5	3.5	7.5	1.5
4.5	11	3.5	5	8	6.5

a. Make a histogram of the data.

b. Were there more students on the Internet for fewer than 3 hours or more than 6 hours?

For Exercises 5 and 6, complete each step.

a. Linearize the data according to the given model.

b. Graph the linearized data, and find the linear regression equation.

c. Use the linear model to find a model for the original data. (Lesson 3-5)

5. exponential

x	y
0	11.1
1	40.7
2	149.5
3	548.4
4	2012.1
5	7383.1

6. quadratic

x	y
0	2.0
1	0.9
2	6.0
3	17.3
4	34.8
5	58.5

2 Online Option Take an online self-check Chapter Readiness Quiz at connectED.mcgraw-hill.com.

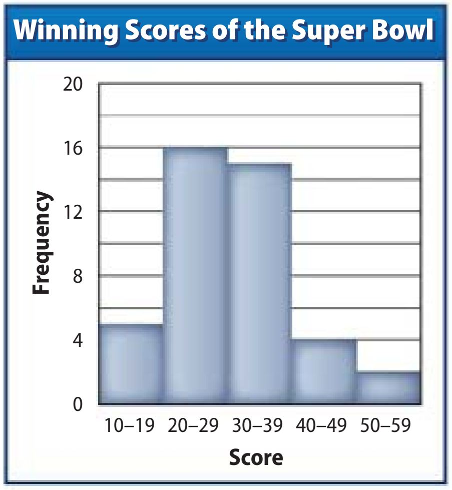
NewVocabulary

English		Español
percentiles	p. 658	percentiles
random variable	p. 664	variable aleatoria
probability distribution	p. 665	distribución probabilística
binomial distribution	p. 669	distribución binomial
normal distribution	p. 674	distribución normal
z-value	p. 676	alor de *z*
standard error of the mean	p. 685	error estándar del media
inferential statistics	p. 696	inferencia estadística
confidence level	p. 696	nivel de confianza
critical values	p. 697	valores críticos
confidence interval	p. 697	intervalo de confianza
t-distribution	p. 699	distribución *t*
hypothesis test	p. 705	prueba de hipótesis
level of significance	p. 706	nivel de significancia
p-value	p. 708	valor *p*
correlation coefficient	p. 713	coeficiente de correlación
regression line	p. 716	recta de regresión
residual	p. 716	residual

ReviewVocabulary

statistics p. P33 estadística the science of collecting, analyzing, interpreting, and presenting data

histogram p. P35 histograma numerical data organized into equal intervals and displayed using bars



LESSON 11-1

Descriptive Statistics

:: Then

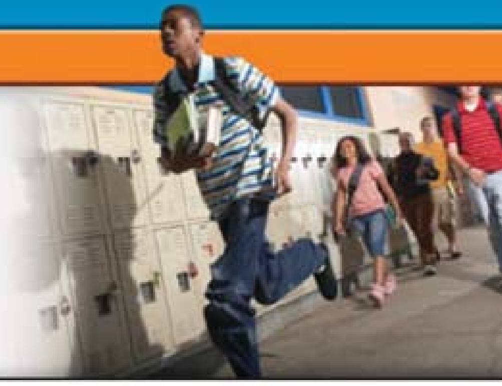
- You found measures of central tendency and standard deviations. (Lesson 0-8)

:: Now

- 1 Identify the shapes of distributions in order to select more appropriate statistics.
- 2 Use measures of position to compare two sets of data.

:: Why?

- A high school newspaper reports that according to a random survey of students, the mean and median number of unexcused tardies received by students last year were 7 and 5, respectively. While both of these values can be used to describe the center of the survey data, only a graph of the data can reveal which measure best represents the typical number of student tardies.



Corbis Super RF/Alamy

NewVocabulary

univariate
negatively skewed distribution
symmetrical distribution
positively skewed distribution
resistant statistic
cluster
bimodal distribution
percentiles
percentile graph

1 Describing Distributions

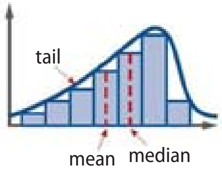
In Lesson 0-8, you described distributions of **univariate** or one-variable data numerically. You did this by calculating and reporting a distribution's

- center using either the mean or median and
- spread or variability using either the standard deviation or five-number summary (quartiles).

To determine which summary statistics you should choose to best describe the center and spread of a data set, you must identify the shape of the distribution. Three common shapes are given below.

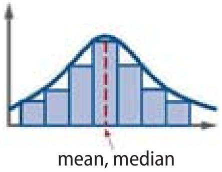
KeyConcept Symmetric and Skewed Distributions

Negative or Left-Skewed Distribution



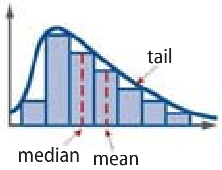
In a **negatively skewed distribution**, the mean is less than the median, the majority of the data is on the right, and the tail extends to the left.

Symmetrical Distribution



In a **symmetrical distribution**, the data are evenly distributed on both sides of the mean. The mean and median are approximately equal.

Positive or Right-Skewed Distribution



In a **positively skewed distribution**, the mean is greater than the median, the majority of the data is on the left, and the tail extends to the right.

When a distribution is reasonably symmetrical, the mean and median are close together. In skewed distributions, however, the mean is located closer to the tail than the median. Outliers, which are extremely high or low values in a data set, will cause the mean to drift even farther toward the tail. The median is less affected by the presence of outliers. For these reasons, the median is called a **resistant statistic** and the mean a *nonresistant statistic*.

Since standard deviation measures the spread of a distribution by how far data values are from the mean, this statistic is also nonresistant to the effects of outliers. This leads to the following guidelines about choosing summary statistics to describe a distribution.

KeyConcept Choosing Summary Statistics

When choosing measures of center and spread to describe a distribution, first examine the shape of the distribution.

- If the distribution is reasonably symmetrical and free of outliers, use the mean and standard deviation.
- If the distribution is skewed or has strong outliers, the five-number summary (minimum, quartile 1, median, quartile 3, maximum) usually provides a better summary of the overall pattern in the data.

When identifying the shape of a distribution, focus on major peaks in the graph instead of minor ups and downs.

Example 1 Skewed Distribution

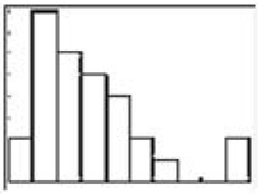
REAL ESTATE The table shows the selling prices for a sample of new homes in a community.

New Home Selling Prices (thousands of dollars)				
248	219	234	250	225
299	205	212	215	245
257	228	221	233	212
220	213	231	212	266
238	249	292	223	235
218	227	209	242	217

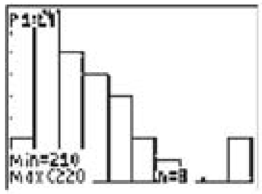
a. Construct a histogram and use it to describe the shape of the distribution.

On a graphing calculator, press STAT Edit and input the data into L1. Then turn on Plot1 under the STAT PLOT menu and choose . Graph the histogram by pressing ZoomStat or by pressing GRAPH and adjusting the window manually.

The graph shown has a single peak. Using the TRACE feature, you can determine that this peak represents selling prices from $210 to $220 thousand.



[200, 300] scl: 10 by [0, 8] scl: 1

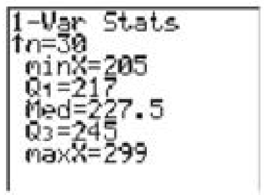


[200, 300] scl: 10 by [0, 8] scl: 1

The graph is positively skewed. Most of the selling prices appear to fall between $210 and $250 thousand, but a few were much higher, so the tail of the distribution trails off to the right.

b. Summarize the center and spread of the data using either the mean and standard deviation or the five-number summary. Justify your choice.

Since the distribution is skewed, use the five-number summary instead of the mean and standard deviation to summarize the center and spread of the data. To display this summary, press STAT , select 1-Var Stats under the CALC submenu, and scroll down.



The five-number summary (minX, Q1, Med, Q3, and maxX) indicates that while the prices range from $205 to $299 thousand, the median selling price was $227.5 thousand and half of the prices were between $217 and $245 thousand.

GuidedPractice

1. LAB GRADES The laboratory grades of all of the students in a biology class are shown.

A. Construct a histogram, and use it to describe the shape of the distribution.

B. Summarize the center and spread of the data using either the mean and standard deviation or the five-number summary. Justify your choice.

Lab Grades (percent)					
72	84	67	80	75	87
86	76	89	91	96	74
68	83	80	76	63	98
92	73	80	88	94	78

TechnologyTip

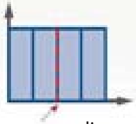
Bin Width On a graphing calculator, each bar is called a *bin*. The bin widths are chosen by the calculator when you use the ZoomStat feature. The bin width can be adjusted by changing the Xscl parameter under WINDOW. A bin width that is too narrow or too wide will affect the apparent shape of a distribution.

WatchOut!

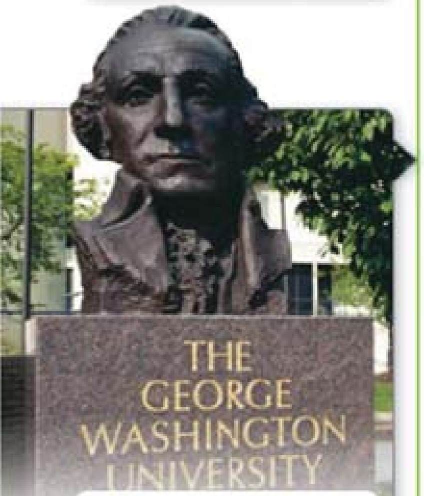
Skew Direction The tail of the distribution indicates in which direction a distribution is skewed, not the peak.

StudyTip

Uniform Distribution In another type of distribution, known as a *uniform distribution*, each value has the same relative frequency, as shown below.



mean, median



Real-WorldLink

In 2008, George Washington University had the highest tuition costs in the U.S. at $37,820 per year. This was roughly 82% of the median annual family income of $46,326.

Source: Forbes Magazine

Distributions of data are not always symmetrical or skewed. Sometimes data will fall into subgroups or **clusters**. If a distribution has a gap in the middle, two separate clusters of data may result. A distribution of data that has two modes, and therefore two peaks, is known as a **bimodal distribution**.

In data that represent a reported preference about a topic, a bimodal distribution can indicate a polarization of opinions. Often, however, a bimodal distribution indicates that the sample data comes from two or more overlapping distributions.

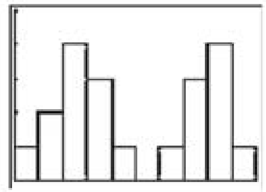
Real-World Example 2 Bimodal Distribution

TUITION **The annual cost of tuition for a sample of 20 colleges at a college fair are shown.**

College Tuition Costs ($)				
32,000	10,100	31,000	11,000	31,500
5500	35,000	10,800	3600	11,500
7400	15,100	18,200	25,600	33,100
36,200	32,000	30,400	14,300	12,400

a. Construct a histogram, and use it to describe the shape of the distribution.

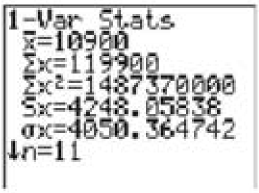
The histogram of the data has not one but two major peaks. Therefore, the distribution is neither symmetrical nor skewed but bimodal. The two separate clusters suggest that two types of colleges are mixed in the data set. It is likely that the 11 colleges with less expensive tuitions are public colleges, and the 9 colleges with more expensive tuitions are private colleges.

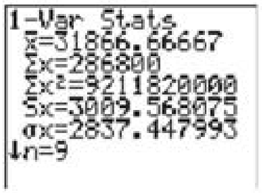


[0, 40,000] scl: 4000 by [0, 6] scl: 1

b. Summarize the center and spread of the data using either the mean and standard deviation or the five-number summary. Justify your choice.

Since the distribution is bimodal, an overall summary of center and spread would give an inaccurate depiction of the data. Instead, summarize the center and spread of each cluster. Since each cluster appears fairly symmetrical, enter each cluster separately and summarize the data using the mean and standard deviation of each cluster.





The mean cost of Cluster 1 is $10,900 with a standard deviation of about $4050, while the mean cost of the Cluster 2 is $31,866 with a standard deviation of about $2837.

GuidedPractice

2. **TRACK** The numbers of minutes that 30 members of a high school cross-country team ran during a practice session are shown.

Practice Session Times (min)									
26	36	31	58	51	29	56	23	61	46
30	50	45	22	64	49	34	42	53	55
41	37	28	54	32	50	59	48	62	39

A. Construct a histogram, and use it to describe the shape of the distribution.

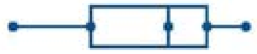
B. Summarize the center and spread of the data using either the mean and standard deviation or the five-number summary. Justify your choice.

Jonathan Hodson/Photolibrary

You can also examine a box-and-whisker plot or *box plot* of a set of data to identify the shape of a distribution. To determine symmetry or skewness from a box plot, you must consider both the position of the line representing the median and the length of each "whisker."

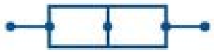
KeyConcept Symmetric and Skewed Box Plots

Negatively Skewed



The left whisker is longer than right whisker, and the line representing the median is closer to Q_3 than to Q_1.

Symmetrical



The whiskers are the same length, and the line representing the median is exactly between Q_1 and Q_3.

Positively Skewed

The right whisker is longer than left whisker, and the line representing the median is closer to Q_1 than to Q_3.

Example 3 Describe a Distribution Using a Box Plot

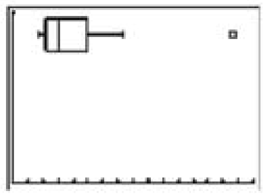
POPULATION The table shows the populations, in thousands of people, during a recent year for fifteen cities in Florida.

Population (Thousands)				
151	95	303	89	186
362	137	109	152	118
102	226	139	736	248

a. Construct a box plot, and use it to describe the shape of the distribution.

Input the data into L1 on a graphing calculator. Then turn on Plot1 under the STAT PLOT menu and choose . Graph the box plot by pressing ZoomStat or by pressing WINDOW and adjusting the window manually.

Since the right whisker is longer than the left whisker and the line representing the median is closer to Q_1 than to Q_3, the distribution is positively skewed. Notice that the distribution has an outlier at 736.



[0, 800] scl: 50 by [0, 1] scl: 0.5

b. Summarize the center and spread of the data using either the mean and standard deviation or the five-number summary. Justify your choice.

Since the distribution is skewed, use the five-number summary. This summary indicates that while the populations ranged from 89,000 to 736,000, the median population was 151,000. Populations in the middle half of the data varied by 248,000 – 109,000 or 139,000 people, which is the interquartile range.

ReviewVocabulary

interquartile range the difference between the upper quartile and lower quartile of a data set (Lesson 0-8)

GuidedPractice

3. TRUCKS The costs on a used car web site for twelve trucks that are the same make, model, and year are shown.

A. Construct a box plot, and use it to describe the shape of the distribution.

B. Summarize the center and spread of the data using either the mean and standard deviation or the five-number summary. Justify your choice.

Used Truck Costs ($)		
9000	8200	9200
7800	8900	8500
6500	7500	7800
8000	6400	5500

2 Measures of Position

The quartiles given by the five-number summary specify the positions of data values within a distribution. For this reason, box plots are most useful for side-by-side comparisons of two or more distributions.

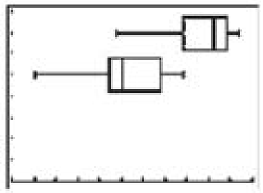
Example 4 Compare Position Using Box Plots

BASKETBALL The number of games won by the Boston Celtics during two different 15-year periods are shown. Construct side-by-side box plots of the data sets. Then use this display to compare the distributions.

1st 15-Year Period				
49	52	59	57	60
60	54	62	59	58
54	48	34	44	56

2nd 15-Year Period				
48	32	35	33	15
36	19	36	35	49
44	36	45	33	24

Input the data into L1 and L2. Then turn on Plot1 and Plot2 under the STAT PLOT menu, choose ⊏⊐, and graph the box plots by pressing ZoomStat or by pressing GRAPH and adjusting the window manually.



[10, 65] scl: 5 by [0, 1] scl: 0.5

Compare Measures of Position

The median number of games won each season in the first 15-year period is greater than those won every season in the second 15-year period. The first quartile for the first 15-year period is approximately equal to the maximum value for the second 15-year period. This means that 75% of the data values for the first 15-year period are greater than any of the values in the second 15-year period. Therefore, we can conclude that the Celtics had significantly more successful seasons during the first 15-year period than during the second 15-year period.

Compare Spreads

The spread of the middle half of the data, represented by the box, is roughly the same in each distribution. Therefore, the variability in the number of games won each season in those 15-year periods was about the same.

GuidedPractice

4. **BASEBALL** The number of home runs hit in Major League Baseball in 1927 and 2007 by the top 20 home run hitters is shown. Construct side-by-side box plots of the data sets. Then use this display to compare the distributions.

1927				
19	10	13	30	16
14	18	14	12	60
30	14	47	14	15
12	20	18	26	17

2007				
40	32	47	31	34
34	33	35	30	46
32	54	31	33	32
36	31	34	50	35

StudyTip

Fractiles Quartiles and percentiles are two types of *fractiles*—numbers that divide an ordered set of data into equal groups. *Deciles* divide a data set into ten equal groups.

In addition to quartiles, you can also use *percentiles* to indicate the relative position of an individual value within a data set. **Percentiles** divide a distribution into 100 equal groups and are symbolized by $P_1, P_2, P_3, \ldots, P_{99}$. The nth percentile or P_n, is the value such that n% of the data are less than P_n and $(100 - n)$% of the data are equal to or greater than P_n. The highest percentile that a data value can be is the 99^{th} percentile.

A **percentile graph** uses the same values as a cumulative relative frequency graph, except that the proportions are instead expressed as percents.

You can use a percentile graph to approximate the percentile rank of a given value for a variable.

Example 5 Construct and Use a Percentile Graph

GPA The table gives the frequency distribution of the GPAs of the 200 students at Ashlyn's high school.

Class Boundaries	f	Class Boundaries	f
2.00–2.25	10	3.00–3.25	36
2.25–2.50	28	3.25–3.50	32
2.50–2.75	30	3.50–3.75	26
2.75–3.00	32	3.75–4.00	6

WatchOut!

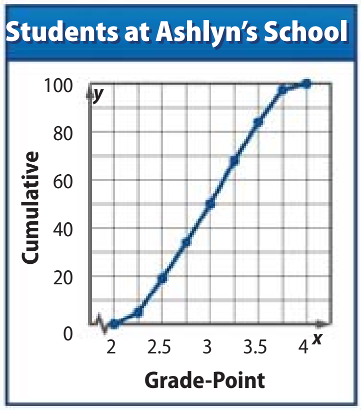
Percentage Versus Percentile

Percentages are not the same as *percentiles*. If a student gets 85 problems correct out of a possible 100, he obtains a percentage score of 85. This does not indicate whether the grade was high or low compared to the rest of the class.

a. Construct a percentile graph of the data.

First, find the cumulative frequencies. Then find the cumulative percentages by expressing the cumulative frequencies as percents. The calculations for the first two classes are shown.

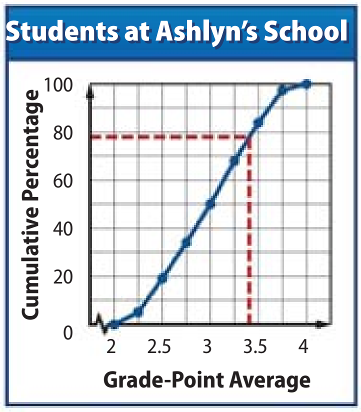
Class Boundaries	f	Cumulative Frequency	Cumulative Percentages
2.00–2.25	10	10	$\frac{10}{200}$ or 5%
2.25–2.50	28	10 + 28 or 38	$\frac{38}{200}$ or 19%
2.50–2.75	30	68	34%
2.75–3.00	32	100	50%
3.00–3.25	36	136	68%
3.25–3.50	32	168	84%
3.50–3.75	26	194	97%
3.75–4.00	6	200	100%



Finally, graph the data with the class boundaries along the x-axis and the cumulative percentages along the y-axis, as shown.

b. Estimate the percentile rank a GPA of 3.4 would have in this distribution, and interpret its meaning.

Find 3.4 on the x-axis and draw a vertical line to the graph. This point on the graph corresponds to approximately the 78th percentile. Therefore, a student with a GPA of 3.4 has a better grade-point average than about 78% of the students at Ashlyn's school.



GuidedPractice

5. HEIGHT The table gives the frequency distribution of the heights of girls in Mr. Lee's precalculus classes.

A. Construct a percentile graph of the data.

B. Estimate the percentile rank a girl with a height of 68 inches would have in this distribution, and interpret its meaning.

Class Boundaries	Frequency (f)
58.5–61.5	11
61.5–64.5	15
64.5–67.5	15
67.5–70.5	12
70.5–73.5	7

WatchOut!

Understanding Percentiles

Saying that a girl's height is at the 75th percentile does *not* mean that her height is 75% of some ideal height. Instead, her height is greater than 75% of all girls in the precalculus class.

Exercises

● = Step-by-Step Solutions begin on page R29.

For Exercises 1–4, complete each step.

a. **Construct a histogram, and use it to describe the shape of the distribution.**

b. **Summarize the center and spread of the data using either the mean and standard deviation or the five-number summary. Justify your choice.** (Examples 1 and 2)

1. **AVIATION** The landing speeds in miles per hour of 20 commercial airplane flights at a certain airport are shown.

Landing Speeds (mph)			
150	157	153	145
155	158	158	162
149	142	138	154
156	161	146	148
158	144	151	152

2. **COMPUTERS** The retail prices of laptop and desktop computers at a certain electronics store are shown.

Computer Prices ($)		
950	1000	975
1150	450	1075
675	1250	540
1025	1180	925
580	950	890

3. **BOWLING** Bowling scores range from 0 to 300. The scores for randomly selected players at a certain bowling alley are shown.

Bowling Scores				
116	81	234	173	75
61	205	92	219	156
134	259	273	53	241
105	190	94	127	235
228	248	271	46	112
99	223	142	217	68

4. **SALARIES** The starting salary for an employee at a certain new company ranges from $20,000 to $90,000. Starting salary depends in part on the employee's years of previous experience and the level of the position for which they were hired. The starting salaries for all the company's new-hires last year are shown.

Salaries (thousands of dollars)			
24	40	34	59
48	52	65	54
68	26	85	32
36	42	33	45
38	89		

For Exercises 5–6, complete each step.

a. **Construct a box plot, and use it to describe the shape of the distribution.**

b. **Summarize the center and spread of the data using either the mean and standard deviation or the five-number summary. Justify your choice.** (Example 3)

5. **VIDEO GAMES** The amount of time that a sample of students at East High School spends playing video games each week is shown.

Time Spent Playing Video Games (hours)					
1.5	2.5	0	4.5	12.5	1
2.5	4	2	8.5	1.5	9
1	0	2	1.5	5.5	2

6. **ACT** The students in Mrs. Calhoun's homeroom class recently took the ACT. The score for each student is shown.

ACT Scores							
32	21	24	35	28	29	28	30
28	25	29	19	24	23	25	22
23	29	27	24	27	29	21	18

For Exercises 7–8, complete each step.

a. **Construct side-by-side box plots of the data sets.**

b. **Use this display to compare the two distributions.** (Example 4)

7. **HYBRID CARS** The fuel efficiency in miles per gallon for 18 hybrid cars manufactured during two recent years are shown.

Year 1								
23	48	31	27	28	35	27	28	24
15	16	28	33	22	16	28	40	24
Year 2								
29	34	25	33	26	35	27	40	27
22	48	29	34	21	24	29	21	34

8. **EARTHQUAKES** The Richter scale magnitudes of 18 earthquakes that occurred in recent years in Alaska and California are shown.

Alaska								
6.6	6.6	6.4	7.2	6.5	6.7	4.8	6.8	6.8
7.8	6.9	7.1	6.6	7.9	6.7	5.3	7.9	7.7
California								
5.4	5.4	5.6	4.4	4.2	4.3	5.2	4.5	4.7
6.6	4.9	7.2	5.2	4.1	6.0	3.0	6.6	3.5

9. MARINE BIOLOGY The table gives the frequency distribution of the weights, in pounds, of 40 adult female sea otters in Washington. (Example 5)

a. Construct a percentile graph of the data.

b. Estimate the percentile rank a weight of 55 pounds would have in this distribution, and interpret its meaning.

Class Boundaries	*f*
40.5–45.5	4
45.5–50.5	5
50.5–55.5	7
55.5–60.5	12
60.5–65.5	9
65.5–70.5	3

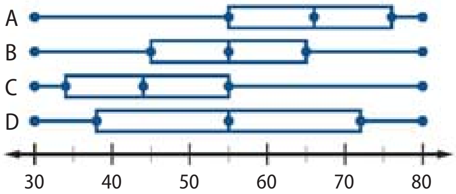
10. RAINFALL The table gives the frequency distribution of the average annual rainfall in inches for all 50 U.S. states. (Example 5)

Class Boundaries	*f*
0–9.5	3
9.5–19.5	8
19.5–29.5	4
29.5–39.5	14
39.5–49.5	16
49.5–59.5	5

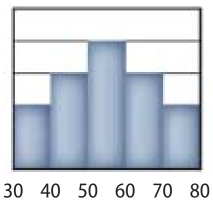
a. Construct a percentile graph of the data.

b. Estimate the percentile rank an average rainfall of 50 inches would have in this distribution, and interpret its meaning.

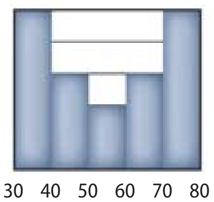
Write the letter of the box plot that corresponds to each of the following histograms.



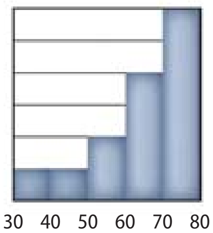
11.



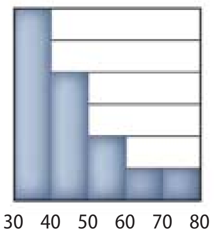
12.



13.



14.



15. ATTENDANCE The average number of New York Yankee home game attendees in thousands of people, per season, from 1979 to 2008 is shown.

Home Game Attendance					
31.7	32.4	30.2	25.2	27.9	22.5
27.5	28.0	30.0	32.7	27.0	24.8
23.0	21.6	29.8	29.7	23.5	27.8
31.9	36.5	40.7	38.0	40.8	42.7
42.8	47.8	50.5	51.9	52.7	53.1

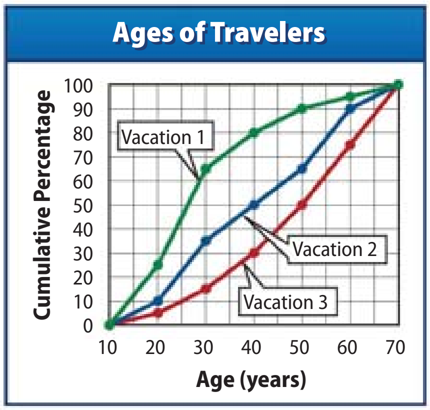
a. Construct a histogram and box plot, and use the graphs to describe the shape of the distribution.

b. Find the average number of people who attended home games during the past 30 years.

c. Which of the graphs would be best to use when estimating the average? Explain your reasoning.

d. Can either of the graphs from part **a** be used to describe any trends in home game attendance during that period? Explain your reasoning.

16. VACATION The percentile graph represents the ages of people who went on three different two-week vacations.



a. Describe the shape of each of the distributions.

b. Which vacation had younger travelers? older travelers? Explain your reasoning.

17. MANUFACTURING The lifetimes, measured in number of charging cycles, for two brands of rechargeable batteries are shown.

Brand A				
998	950	1020	1003	990
942	1115	973	1018	981
1047	1002	997	1110	1003
Brand B				
892	1044	1001	999	903
950	998	993	1002	995
990	1000	1005	997	1004

a. Construct a histogram of each data set.

b. Which of the brands has a greater variation in lifetime?

18. **BASKETBALL** The heights in inches for the players on the U.S. men's and women's national basketball teams during the 2008 Olympics are shown.

Men's Heights					
81	76	80	75	78	76
78	83	82	72	81	80
Women's Heights					
69	73	69	68	73	77
72	74	72	78	71	76

a. Construct a percentile graph of the data.

b. Estimate the percentile ranks that a male and a female player with a height of 75 inches would have in each distribution. Interpret their meaning.

c. Suppose the 78-inch-tall women's player is replaced with a 74-inch-tall player. What percentile rank would the new player have in the corresponding distribution?

Another measure of center known as the *midquartile* is given by $\frac{Q_1 + Q_3}{2}$. Find Q_1, Q_2, Q_3, and the midquartile for each set of data.

19.

0.12	0.25	0.19	0.38	0.28	0.16
0.41	0.29	0.32	0.11	0.04	0.25
0.29	0.07	0.26	0.09	0.31	0.23

20.

112	101	138	200	176	199
105	127	146	128	116	154
167	202	191	143	205	130

21. **ENERGY** Petroleum consumption from 1988 to 2007 for the United States and North America is shown.

United States (thousands of barrels/day)				
16,700	17,300	17,300	17,000	16,700
17,000	17,200	17,700	17,700	18,300
18,600	18,900	19,500	19,700	19,600
19,800	20,000	20,700	20,800	20,700
North America (thousands of barrels/day)				
19,900	20,600	20,800	20,000	20,200
20,600	20,800	21,400	21,300	22,000
22,400	22,800	23,500	23,800	23,700
23,800	24,200	25,000	25,200	25,000

a. Construct side-by-side box plots and histograms.

b. Compare the average petroleum consumption for the U.S. and North America.

c. Which of the graphs is easier to use when comparing measures of center and spread?

d. On average, what percent of petroleum consumption in North America can be attributed to the U.S.? Round to the nearest percent.

22. **MULTIPLE REPRESENTATIONS** In this problem, you will investigate how a linear transformation affects the shape, center, and spread of a distribution of data. Consider the table shown.

52	37	59	31	45
23	48	42	65	39
40	53	14	49	56
68	32	77	44	28

a. GRAPHICAL Construct a histogram and use it to describe the shape of the distribution.

b. NUMERICAL Find the mean and standard deviation of the data set.

c. TABULAR Perform each of the following linear transformations of the form $X' = a + bX$, where X is the initial data value and X' is the transformed data value. Record each set of transformed data values (**i–iii**) in a separate table.

i. $a = 3, b = 5$ **ii.** $a = 10, b = 1$ **iii.** $a = 0, b = 5$

d. GRAPHICAL Repeat parts **a** and **b** for each set of transformed data values that you found in part **c**. Adjust the bin width for each appropriately.

e. VERBAL Describe how a linear transformation affects the shape, center, and spread of a distribution of data.

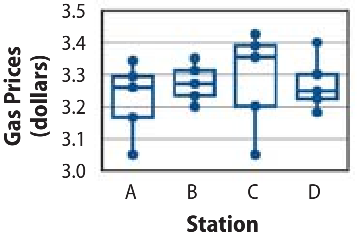
f. ANALYTICAL If every value in a data set is multiplied by a constant c, what will happen to the mean and standard deviation of the distribution?

H.O.T. Problems Use Higher-Order Thinking Skills

23. **WRITING IN MATH** Explain why using the range can be an ineffective method for measuring the spread of a distribution of data.

24. **CHALLENGE** Suppose 20% of a data set lies between 35 and 55. If 10 is added to each value in the set and then each result is doubled, what values will 20% of the resulting data lie between?

REASONING **The gas prices, in dollars, at four gas stations over a period of one month are shown.**



25. Which of the stations has the greatest variation in gas prices? the least variation? Explain your reasoning.

26. Which of the distributions is positively skewed? negatively skewed? symmetrical? Explain your reasoning.

27. **WRITING IN MATH** Why is the median less affected by outliers than the mean? Justify your answer.

Spiral Review

Write each complex number in exponential form. (Lesson 10-6)

28. $\sqrt{3} + \sqrt{3}i$

29. $\sqrt{5} - \sqrt{5}i$

30. $\sqrt{2} - \sqrt{6}i$

Use the Binomial Theorem to expand each binomial. (Lesson 10-5)

31. $(3a + 4b)^5$

32. $(5c - 2d)^4$

33. $(-2x + 4y)^6$

Find the sum of each geometric series described. (Lesson 10-3)

34. first five terms of $\frac{5}{3} + 5 + 15 + \cdots$

35. first six terms of $65 + 13 + 2.6 + \cdots$

36. first ten terms of $1 - \frac{3}{2} + \frac{9}{4} - \cdots$

Find the angle θ between vectors u and v. (Lesson 8-5)

37. $\mathbf{u} = 4\mathbf{i} - 2\mathbf{j} + 9\mathbf{k}$, $\mathbf{v} = 3\mathbf{i} + 7\mathbf{j} - 10\mathbf{k}$

38. $\mathbf{u} = \langle -7, 4, 2 \rangle$, $\mathbf{v} = \langle 9, -5, 1 \rangle$

39. $\mathbf{u} = \langle 4, 4, -6 \rangle$, $\mathbf{v} = \langle 8, -5, 2 \rangle$

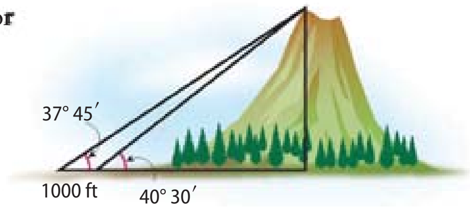
Graph the ellipse given by each equation. (Lesson 7-2)

40. $\frac{x^2}{4} + \frac{(y-3)^2}{25} = 1$

41. $\frac{(x+6)^2}{16} + \frac{(y-5)^2}{9} = 1$

42. $\frac{(x-2)^2}{28} + \frac{y^2}{8} = 1$

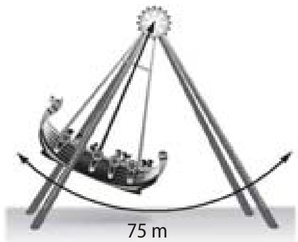
43. **SURVEYING** To determine the new height of a volcano after an eruption, a surveyor measured the angle of elevation to the top of the volcano to be 37° 45′. The surveyor then moved 1000 feet closer to the volcano and measured the angle of elevation to be 40° 30′. Determine the new height of the volcano. (Lesson 4-1)



Skills Review for Standardized Tests

44. **REVIEW** An amusement park ride operates like the bob of a pendulum. On its longest swing, the ship travels through an arc 75 meters long. Each successive swing is two-fifths the length of the preceding swing. What is the total distance the ship will travel from the beginning of its longest swing if the ride is allowed to continue without intervention?

A 75 m
B 125 m
C 150 m
D 187.5 m



45. **REVIEW** The value of a certain car depreciated at a constant rate. If the initial value was $25,000 and the car was worth $8192 after five years, find the annual rate of depreciation.

F 10%
G 20%
H 30%
J 40%

46. **SAT/ACT** The values of each house in a city are collected and analyzed. Which descriptive statistic will best describe the data?

A mean
B median
C mode
D range
E standard deviation

47. The table shows the frequency distribution of scores on the state driving test at a particular center on a given day. Estimate the percentile rank of someone who scored a 72 that day.

F 27%
G 30%
H 34%
J 72%

Class Boundaries	Frequency f
0–65.5	12
65.5–70.5	3
70.5–75.5	4
75.5–80.5	1
80.5–85.5	9
85.5–90.5	13
90.5–95.5	8
95.5–100	6

LESSON 11-2 Probability Distributions

Then

- You found probabilities of events involving combinations. (Lesson 0-7)

Now

1. Construct a probability distribution, and calculate its summary statistics.
2. Construct and use a binomial distribution, and calculate its summary statistics.

Why?

- Car insurance companies use statistics to measure the risk associated with particular events, such as collisions. Using data about what has happened in the past, they assign probabilities to all possible outcomes relating to the event and calculate statistics based on how these probabilities are distributed. With these statistics, they can predict the likelihood of certain outcomes and make decisions accordingly.

NewVocabulary

random variable
discrete random variable
continuous random variable
probability distribution
expected value
binomial experiment
binomial distribution
binomial probability distribution function

1 Probability Distributions

In the previous lesson, you used descriptive statistics to analyze a *variable*, a characteristic of a population. In that lesson, the values the variable could take on were determined by collecting data. In this lesson, you will consider variables with values that are determined by chance.

A **random variable** X represents a numerical value assigned to an outcome of a probability experiment. There are two types of random variables: discrete and continuous.

KeyConcept Discrete and Continuous Random Variables

A **discrete random variable** can take on a finite or countable number of possible values.

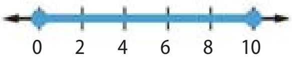
Example

Number of Car Accidents Per Year, X

A **continuous random variable** can take on an infinite number of possible values within a specified interval.

Example

Number of Miles Driven, X



Since different statistical techniques are used to analyze these two types of random variables, it is important to be able to distinguish between them. To correctly classify a random variable, consider whether X represents counted or measured data.

Example 1 Classify Random Variables as Discrete or Continuous

Classify each random variable X as *discrete* or *continuous*. Explain your reasoning.

a. **X represents the weight of the cereal in a 15-ounce box of cereal chosen at random from those on an assembly line.**

The weight of the cereal could be any weight between 0 and 15 ounces. Therefore, X is a continuous random variable.

b. **X represents the number of cars in a school parking lot chosen at a random time during the school day.**

The number of cars in the parking lot is countable. There could be 0, 1, 2, 3, or some other whole number of cars. Therefore, X is a discrete random variable.

GuidedPractice

1A. X represents the time it takes to serve a fast-food restaurant customer chosen at random.

1B. X represents the attendance at a randomly selected monthly school board meeting.

The sample space for the familiar theoretical probability experiment of tossing two coins is {TT, TH, HT, HH}. If X is the random variable for the number of heads, then X can assume the value 0, 1, or 2. From the sample space, you can find the theoretical probability of getting no, one, or two heads.

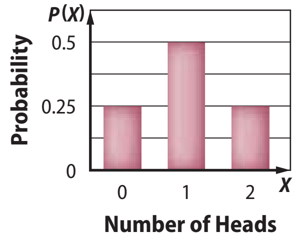
$P(0) = \frac{1}{4}$ $\quad P(1) = \frac{1}{2}$ $\quad P(2) = \frac{1}{4}$

ReadingMath

Probabilities of Random Variables The notation $P(1)$ is read *the probability that the random variable X is equal to 1.*

The table below and the graph at the right show the *probability distribution* of X.

Number of heads, X	0	1	2
Probability, $P(X)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$



KeyConcept Probability Distribution

A **probability distribution** of a random variable X is a table, equation, or graph that links each possible value of X with its probability of occurring. These probabilities are determined theoretically or by observation.

A probability distribution must satisfy the following conditions.

- The probability of each value of X must be between 0 and 1. That is, $0 \leq P(X) \leq 1$.
- The sum of all the probabilities for all the values of X must equal 1. That is, $\sum P(X) = 1$.

StudyTip

Continuous Distributions This lesson focuses on discrete random variables. You will study continuous probability distributions in Lesson 11–3.

To construct a discrete probability distribution using observed instead of theoretical data, use the frequency of each observed value to compute its probability.

Example 2 Construct a Probability Distribution

TEACHER EVALUATION On a teacher evaluation form, students were asked to rate the teacher's explanations of the subject matter using a score from 1 to 5, where 1 was too simplified and 5 was too technical. Use the frequency distribution shown to construct and graph a probability distribution for the random variable X.

Score, X	Frequency
1	1
2	8
3	20
4	16
5	5

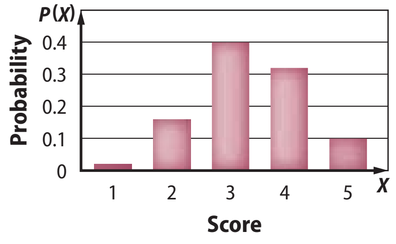
To find the probability that X takes on each value, divide the frequency of each value by the total number of students rating this teacher, which is $1 + 8 + 20 + 16 + 5$ or 50.

$P(1) = \frac{1}{50}$ or 0.02 $\quad P(2) = \frac{8}{50}$ or 0.16 $\quad P(3) = \frac{20}{50}$ or 0.40

$P(4) = \frac{16}{50}$ or 0.32 $\quad P(5) = \frac{5}{50}$ or 0.10

The probability distribution of X is shown below, and its graph is shown at the right.

Score, X	1	2	3	4	5
$P(X)$	0.02	0.16	0.40	0.32	0.10



CHECK Note that all of the probabilities in the table are between 0 and 1 and that $\sum P(X) = 0.02 + 0.16 + 0.4 + 0.32 + 0.1$ or 1. ✓

GuidedPractice

2. **CAR SALES** A car salesperson tracked the number of cars she sold each day during a 30-day period. Use the frequency distribution of the results to construct and graph a probability distribution for the random variable X, rounding each probability to the nearest hundredth.

Cars Sold, X	0	1	2	3
Frequency	20	7	2	1

To compute the mean of a probability distribution, we must use a formula different from that used to compute the mean of a population. To understand why, consider computing the mean of the number of heads X resulting from an infinite number of two-coin tosses. We cannot compute the mean using $\mu = \frac{\sum X}{N}$, since N would be infinite. However, the probability distribution of X tells us what fraction of those tosses we would *expect* to have a value of 0, 1, or 2.

Number of Heads after Two Coin Tosses

$$\{TT,\ TT,\ \ldots,\ TT,\ TT,\quad HT,\ HT,\ \ldots,\ HT,\ HT,\ TH,\ TH,\ \ldots,\ TH,\ TH,\quad HH,\ HH,\ \ldots,\ HH,\ HH\}$$
$$\{\underbrace{0,\ 0\ \ldots,\ 0,\ 0,}_{\frac{1}{4}}\quad \underbrace{1,\ 1,\ \ldots,\ 1,\ 1,\ 1,\ 1,\ \ldots,\ 1,\ 1,}_{\frac{1}{2}}\quad \underbrace{2,\ 2,\ \ldots,\ 2,\ 2}_{\frac{1}{4}}\}$$

Therefore, we would expect that on average the number of heads for many or an infinite number of tosses would be $\frac{1}{4} \cdot 0 + \frac{1}{2} \cdot 1 + \frac{1}{4} \cdot 2$ or 1. This method for finding the mean of a probability distribution is summarized below.

KeyConcept Mean of a Probability Distribution

Words To find the mean of a probability distribution of X, multiply each value of X by its probability and find the sum of the products.

Symbols The mean of a random variable X is given by $\mu = \sum[X \cdot P(X)]$, where $X_1, X_2, \ldots, X_n$ are the values of X and $P(X_1), P(X_2), \ldots, P(X_n)$ are the corresponding probabilities.

Example 3 Mean of a Probability Distribution

TEACHER EVALUATION **The table shows the probability distribution for the teacher evaluation question from Example 2. Find the mean score to the nearest hundredth, and interpret its meaning in the context of the problem situation.**

Score, X	$P(X)$
1	0.02
2	0.16
3	0.40
4	0.32
5	0.10

Multiply each score by its probability, and find the sum of these products. Organize your calculations by extending the table.

Score, X	$P(X)$	$X \cdot P(X)$
1	0.02	$1 \cdot 0.02 = 0.02$
2	0.16	$2 \cdot 0.16 = 0.32$
3	0.40	$3 \cdot 0.40 = 1.20$
4	0.32	$4 \cdot 0.32 = 1.28$
5	0.10	$5 \cdot 0.10 = 0.50$
		$\sum[X \cdot P(X)]$ or 3.3

Therefore, the mean μ of this probability distribution is about 3.3.

Since a score of 3 indicates that the teacher's explanations were neither too simplified nor too technical, a mean of 3.3 indicates that on average, students felt that this teacher's explanations were appropriate but leaned slightly towards being too technical.

StudyTip

Rounding Rule The mean, as well as the variance and standard deviation, discussed on the next page, should be rounded to one decimal place more than that of an actual value that X can assume.

GuidedPractice

3. **CAR SALES** Find the mean of the probability distribution that you constructed in Guided Practice 2 and interpret its meaning in the context of the problem situation.

The variance formula used for population distributions can also not be used to calculate the variance or standard deviation of a probability distribution, because the value of N would be infinite. Instead, the following formulas are used to find the spread of a probability distribution.

KeyConcept Variance and Standard Deviation of Probability Distribution

Words To find the variance of a probability distribution of X, subtract the mean of the probability distribution from each value of X and square the difference. Then multiply each difference by its corresponding probability and find the sum of the products. The standard deviation is the square root of the variance.

Symbols The variance of a random variable X is given by $\sigma^2 = \sum[(X - \mu)^2 \cdot P(X)]$, and the standard deviation is given by $\sigma = \sqrt{\sigma^2}$.

StudyTip

Alternate Formula A mathematically equivalent formula for the variance of a probability distribution that can significantly simplify the calculation of this statistic is $\sigma^2 = \sum[(X^2 \cdot P(X)] - \mu^2$.

Example 4 Variance and Standard Deviation of a Probability Distribution

TEACHER EVALUATION **Find the variance and standard deviation of the probability distribution for the teacher evaluation question from Example 2 to the nearest hundredth.**

Score, X	$P(X)$
1	0.02
2	0.16
3	0.40
4	0.32
5	0.10

Subtract each value of X from the mean found in Example 3, 3.32 and square the difference. Then multiply each difference by its corresponding probability and find the sum of the products.

Score, X	$P(X)$	$(X - \mu)^2$	$(X - \mu)^2 \cdot P(X)$
1	0.02	$(1 - 3.32)^2 \approx 5.38$	$5.38 \cdot 0.02 \approx 0.1076$
2	0.16	$(2 - 3.32)^2 \approx 1.74$	$1.74 \cdot 0.16 \approx 0.2788$
3	0.40	$(3 - 3.32)^2 \approx 0.10$	$0.10 \cdot 0.40 \approx 0.0410$
4	0.32	$(4 - 3.32)^2 \approx 0.46$	$0.46 \cdot 0.32 \approx 0.1480$
5	0.10	$(5 - 3.32)^2 \approx 2.82$	$2.82 \cdot 0.10 \approx 0.2822$
			$\sum[(X - \mu)^2 \cdot P(X)] = 0.8576$

The variance σ^2 is about 0.86, and the standard deviation is $\sqrt{0.8576}$ or about 0.93.

GuidedPractice

4. **CAR SALES** Find the variance and standard deviation of the probability distribution that you constructed in Guided Practice 2 to the nearest hundredth.

The **expected value** $E(X)$ of a random variable for a probability distribution is equal to the mean of the random variable. That is, $E(X) = \mu = \sum[X \cdot P(X)]$.

Example 5 Find an Expected Value

FUNDRAISERS **At a raffle, 500 tickets are sold at \$1 each for three prizes of \$100, \$50, and \$10. What is the expected value of your net gain if you buy a ticket?**

Construct a probability distribution for the possible net gains. Then find the expected value. The net gain for each prize is the value of the prize minus the cost of the tickets purchased.

Gain, X	\$100 − 1 or \$99	\$50 − 1 or \$49	\$10 − 1 or \$9	\$0 − 1 or −\$1
Probability, $P(X)$	$\frac{1}{500}$ or 0.002	$\frac{1}{500}$ or 0.002	$\frac{1}{500}$ or 0.002	$\frac{497}{500}$ or 0.994

$$E(X) = \sum[X \cdot P(X)]$$
$$= (99 \cdot 0.002) + (49 \cdot 0.002) + (9 \cdot 0.002) + (-1 \cdot 0.994) \text{ or about } -\$0.68$$

This expected value means that the average loss for someone purchasing a ticket is \$0.68.

WatchOut!

Misinterpreting Expected Value An expected value such as that calculated in Example 5 is *not* an indication of how much a person might win or lose. In Example 5, a person can lose only \$1 for each ticket purchase and can win only \$100, \$50, or \$10.

GuidedPractice

5. **WATER PARK** A water park makes $350,000 when the weather is normal and loses $80,000 per season when there are more bad weather days than normal. If the probability of having more bad weather days than normal this season is 35%, find the park's expected profit.

2 Binomial Distribution

Many probability experiments can be reduced to one involving only two outcomes: success or failure. For example, a multiple-choice question with five answer choices can be classified as simply correct or incorrect, or a medical treatment can be classified as effective or ineffective. Such experiments have been reduced to *binomial* experiments.

KeyConcept Binomial Experiment

A **binomial experiment** is a probability experiment that satisfies the following conditions.

- The experiment is repeated for a fixed number of independent trials n.
- Each trial has only two possible outcomes, success S or failure F.
- The probability of success $P(S)$ or p is the same in every trial. The probability of failure $P(F)$ or q is $1 - p$.
- The random variable X represents the number of successes in n trials.

Real-WorldLink

One in five American teens ages 12 years and older owns a portable MP3 player. More than one in twenty teens own more than one player.

Source: Digital Trends

Real-World Example 6 Identify a Binomial Experiment

Determine whether each experiment is a binomial experiment or can be reduced to a binomial experiment. If it can be presented as a binomial experiment, state the values of n, p, and q. Then list all possible values of the random variable. If it is not, explain why not.

a. The results of a school survey indicate that 68% of students own an MP3 player. Six students are randomly selected and asked if they own an MP3 player. The random variable represents the number of students who say that they do own an MP3 player.

The experiment satisfies the conditions of a binomial experiment.

- Each student selected represents one trial, and the selection of each of the six students is independent of the others.
- There are only two possible outcomes: either the student owns an MP3 player S or does not own an MP3 player F.
- The probability of success is the same for each student selected, $P(S) = 0.68$.

In this experiment, $n = 6$ and $p = P(S)$ or 0.68. The probability of failure is $q = 1 - p$, so $q = 1 - 0.68$ or 0.32. X represents the number of students who own an MP3 player out of those selected, so $X = 0, 1, 2, 3, 4, 5,$ or 6.

b. Five cards are drawn at random from a deck to make a hand for a card game. The random variable represents the number of spades.

In this experiment, each card selected represents one trial. The probability of drawing a spade for the first card is $\frac{13}{52}$ or $\frac{1}{4}$. However, since this card is kept to make a player's hand, the trials are not independent, and the probability of success for each draw will not be the same. Therefore, this experiment cannot be reduced to a binomial experiment.

GuidedPractice

6A. The results of a survey indicate that 61% of students like the new school uniforms and 24% do not. Twenty students are randomly selected and asked if they like the uniforms. The random variable represents the number who say that they do like the uniforms.

6B. You complete a test by randomly guessing the answers to 20 multiple-choice questions that each have 4 answer choices, only one of which is correct. The random variable represents the number of correct answers.

StudyTip

Look Back Refer to Lesson 10–5 to review binomial expansion and the Binomial Theorem.

The distribution of the outcomes of a binomial experiment and their corresponding probabilities is called a **binomial distribution.** The probabilities in this distribution can be calculated using the following formula, which represents the $p^x q^{n-x}$ term in the binomial expansion of $(p + q)^n$.

KeyConcept Binomial Probability Formula

The probability of X successes in n independent trials of a binomial experiment is

$$P(X) = {}_nC_x\, p^x q^{n-x} = \frac{n!}{(n-x)!\,x!}\, p^x q^{n-x},$$

where p is the probability of success and q is the probability of failure for an individual trial.

Notice that this formula represents a discrete function of the random variable X, known as the **binomial probability distribution function**.

Example 7 Binomial Distributions

EXERCISE In a recent poll, 35% of teenagers said they exercise regularly. Five teenagers chosen at random are asked if they exercise regularly. Construct and graph a binomial distribution for the random variable X, which represents the number of teenagers who said yes. Then find the probability that at least three of these teenagers said yes.

This is a binomial experiment in which $n = 5$, $p = 0.35$, $q = 1 - 0.35$ or 0.65. Use a calculator to compute the probability of each possible value for X using the Binomial Probability Formula.

$P(0) = {}_5C_0 \cdot 0.35^0 \cdot 0.65^5 \approx 0.116$

$P(1) = {}_5C_1 \cdot 0.35^1 \cdot 0.65^4 \approx 0.312$

$P(2) = {}_5C_2 \cdot 0.35^2 \cdot 0.65^3 \approx 0.336$

$P(3) = {}_5C_3 \cdot 0.35^3 \cdot 0.65^2 \approx 0.181$

$P(4) = {}_5C_4 \cdot 0.35^4 \cdot 0.65^1 \approx 0.049$

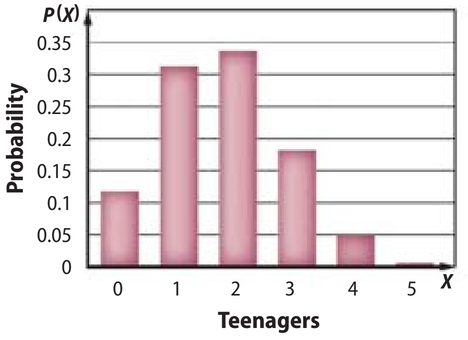
$P(5) = {}_5C_5 \cdot 0.35^5 \cdot 0.65^0 \approx 0.005$

TechnologyTip

Binomial Probability To calculate each binomial probability on a graphing calculator, use binompdf(n, p, x) under the DISTR menu.

The probability distribution of X and its graph are shown below.

X	$P(X)$
0	0.116
1	0.312
2	0.336
3	0.181
4	0.049
5	0.005



To find the probability that *at least* three of the teenagers exercise regularly, find the sum of $P(3)$, $P(4)$, and $P(5)$.

$P(X \geq 3) = P(3) + P(4) + P(5)$	P(at least three)
$= 0.181 + 0.049 + 0.005$	$P(3) = 0.181$, $P(4) = 0.049$, and $P(5) = 0.005$
$= 0.235$ or 23.5%	Simplify.

GuidedPractice

7. **CLASSES** In a certain high school graduating class, 48% of the students took a world language during their senior year. Seven students chosen at random are asked if they took a world language during their final year. Construct and graph a probability distribution for the random variable X, which represents the number of students who said yes. Then find the probability that fewer than 4 of these students said yes.

Use the following formulas to find the mean, variance, and standard deviation of a binomial distribution.

KeyConcept Mean and Standard Deviation of a Binomial Distribution

The mean, variance, and standard deviation of a random variable X that has a binomial distribution are given by the following formulas.

Mean	$\mu = np$
Variance	$\sigma^2 = npq$
Standard Deviation	$\sigma = \sqrt{\sigma^2}$ or $\sqrt{npq}$

These formulas are simpler than, but algebraically equivalent to, the formulas that you used to find the mean, variance, and standard deviation of probability distributions.

Comp: Space here s/b 1p5 b/4 PG 8/24

Real-WorldLink

According to a recent Gallup poll, approximately 58% of U.S. teens are classified as highly active. Activities in which the teens said that they regularly participate include basketball, running, jogging, biking, and swimming.

Source: The Gallup Poll

Real-World Example 8 Mean and Standard Deviation of a Binomial Distribution

EXERCISE The table shows the binomial distribution in Example 7. Find the mean, variance, and standard deviation of this distribution. Interpret the mean in the context of the problem situation.

X	0	1	2	3	4	5
$P(X)$	0.116	0.312	0.336	0.181	0.049	0.005

Method 1 Use the formulas for the mean, variance, and standard deviation of a probability distribution.

$$\mu = \sum[X \cdot P(X)]$$
$$= 0(0.116) + 1(0.312) + 2(0.336) + 3(0.181) + 4(0.049) + 5(0.005)$$
$$= 1.748$$

$$\sigma^2 = \sum[(X - \mu)^2 \cdot P(X)]$$
$$= (0 - 1.748)^2 \cdot 0.116 + (1 - 1.748)^2 \cdot 0.312 + (2 - 1.748)^2 \cdot 0.336 + (3 - 1.748)^2 \cdot 0.181 + (4 - 1.748)^2 \cdot 0.049 + (5 - 1.748)^2 \cdot 0.005$$
$$\approx 1.1354$$

$$\sigma = \sqrt{\sigma^2}$$
$$= \sqrt{1.1354} \text{ or about } 1.0656$$

Method 2 Use the formulas for the mean, variance, and standard deviation of a binomial probability distribution. In this binomial experiment, $n = 5$, $p = 0.35$, and $q = 0.65$.

$$\mu = np$$
$$= 5(0.35) \text{ or } 1.75$$

$$\sigma^2 = npq$$
$$= 5(0.35)(0.65) \text{ or } 1.1375$$

$$\sigma = \sqrt{\sigma^2}$$
$$= \sqrt{1.1375} \text{ or about } 1.0665$$

Both methods give approximately the same results. Therefore, the mean of the distribution is about 1.8 or 2, which means that on average about 2 out of the 5 students would say that they exercise regularly. The variance and standard deviation of the distribution are both about 1.1.

GuidedPractice

8. **CLASSES** Find the mean, variance, and standard deviation of the distribution that you constructed in Guided Practice 7. Interpret the mean in the context of the problem situation.

Exercises

● = Step-by-Step Solutions begin on page R29.

Classify each random variable *X* as *discrete* or *continuous*. Explain your reasoning. (Example 1)

1. X represents the number of text messages sent by a randomly chosen student during a given day.
2. X represents the time it takes a randomly selected student to complete a physics test.
3. X represents the weight of a chocolate chip cookie selected at random in the school cafeteria.
4. X represents the number of CDs owned by a student chosen at random during a given day.
5. X represents the number of votes received by a candidate selected at random for a particular election.
6. X represents the weight of a wrestler selected at random on a given day.

Construct and graph a probability distribution for each random variable *X*. Find and interpret the mean in the context of the given situation. Then find the variance and standard deviation. (Examples 2–4)

7. **MUSIC** Students were asked how many MP3 players they own.

Players, X	Frequency
0	9
1	17
2	9
3	5
4	2

8. **AMUSEMENT** There were 20 participants in a pie eating contest at a county fair.

Pies Eaten, X	Frequency
1	1
2	5
3	9
4	3
5	2

9. **BREAKFAST** A sample of high school students was asked how many days they ate breakfast last week.

Days, X	Frequency
0	5
1	3
2	17
3	27
4	6
5	19
6	18
7	65

10. **HEALTH** Patients at a dentist's office were asked how many times a week they floss their teeth.

Flosses, X	Frequency
1	9
2	15
3	5
4	2
5	1
6	0
7	1

11. **CAR INSURANCE** A car insurance policy that costs $300 will pay $25,000 if the car is stolen and not recovered. If the probability of a car being stolen is $p = 0.0002$, what is the expected value of the profit (or loss) to the insurance company for this policy? (Example 5)

12. **FUNDRAISERS** A school hosts an annual fundraiser where raffle tickets are sold for baked goods, the values of which are indicated below. Suppose 100 tickets were sold for a drawing for each of the four cakes.

What is the expected value of a participant's net gain if he or she buys a ticket for $1? (Example 5)

Determine whether each experiment is a binomial experiment or can be reduced to a binomial experiment. If it can be presented as a binomial experiment, state the values of n, p, and q. Then list all possible values of the random variable. If it is not, explain why not. (Example 6)

13. You survey 25 students to find out how many are left-handed. The random variable represents the number of left-handed people.
14. You survey 200 people to see if they watch Monday Night Football. The random variable represents the number who watched Monday Night Football.
15. You roll a die 10 times to see if a 5 appears. The random variable represents the number of 5s.
16. You toss a coin 20 times to see how many tails occur. The random variable represents the number of tails.
17. You ask 15 people how old they are. The random variable represents their age.
18. You survey 40 students to find out whether they passed their driving test. The random variable represents the number that passed .
19. You select 10 cards from a deck without replacement. The random variable represents the number of hearts.

Construct and graph a binomial distribution for each random variable. Find and interpret the mean in the context of the given situation. Then find the variance and standard deviation. (Examples 7 and 8)

20. In a recent poll, 89% of Americans order toppings on their pizza. Five teenagers chosen at random are asked if they order toppings.

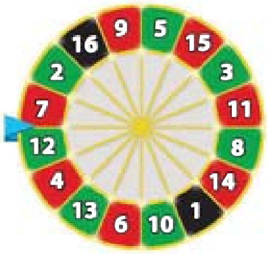
21. In Eureka, California, 21% of the days are sunny. Consider the number of sunny days in February.

22. According to a survey, 26% of a company's employees have surfed the Internet at work. Ten co-workers were selected at random and asked if they have surfed the Internet at work.

23. A high school newspaper reported that 65% of students wear their seatbelts while driving. Eight students chosen at random are asked if they wear seatbelts.

24. According to a recent survey, 41% of high school students own a car. Seven students chosen at random are asked if they own a car.

25. **GAME SHOWS** The prize wheel on a game show has 16 numbers. During one turn, a bet is made and the wheel is spun.



The payoffs for a \$5 bet are shown. If a player bets \$5, find the expected value for each.

Bet	Payoff	Bet	Payoff
red	\$10	1	\$50
green	\$10	16	\$50
1–4	\$15	even and red	\$25
5–8	\$15	odd and green	\$25
9–12	\$15	1 or 16	\$30
even	\$7	odd	\$7

a. green
b. even and red
c. odd
d. 1 or 16
e. 1

26. **VOLUNTEERING** In a recent poll, 62% of Americans said that they had donated their time volunteering for a charity in the past year. If a random sample of 10 Americans is selected, find each of the following probabilities.

a. Exactly 6 people donated their time to a charity.
b. At least 5 people donated their time to a charity.
c. At most 3 people donated their time to a charity.
d. More than 8 people donated their time to a charity.

27. **MULTIPLE REPRESENTATIONS** In this problem, you will investigate the shape of a binomial distribution.

a. **GRAPHICAL** Construct and graph the binomial distribution that corresponds to each of the following experiments.

i. $n = 6, p = 0.5$ **ii.** $n = 6, p = 0.3$
iii. $n = 6, p = 0.7$ **iv.** $n = 8, p = 0.5$
v. $n = 10, p = 0.5$

b. **VERBAL** Describe the shape of each of the distributions you found in part **a.**

c. **ANALYTICAL** Make a conjecture regarding the shape of a distribution with each of the following probabilities of success: $p < 0.5$, $p = 0.5$, and $p > 0.5$.

d. **ANALYTICAL** What happens to the spread of a binomial distribution as n increases?

H.O.T. Problems Use Higher-Order Thinking Skills

28. **PROOF** Use the distribution below to prove that $\mu = np$ and $\sigma^2 = npq$ for a binomial distribution, given $\mu = \sum[X \cdot P(X)]$ and $\sigma^2 = \sum[(X - \mu)^2 \cdot P(X)]$ for a probability distribution.

X	$P(X)$
0	$1 - p$
1	p

29. **REASONING** Suppose a coin is tossed ten times and lands on heads each time. Will the probability of the coin landing on tails increase during the next toss? Explain your reasoning.

30. **OPEN ENDED** A probability distribution in which all of the values of the random variable occur with equal probability is called a *uniform probability distribution*. Describe an example of an experiment that would produce a uniform distribution. Then find the theoretical probabilities that would result from this experiment. Include a table and graph of the distribution.

REASONING **Determine whether each of the following statements is *true* or *false*. Explain your reasoning.**

31. The probabilities associated with rolling two dice are determined theoretically.

32. The mean of a random variable is always a possible outcome of the experiment.

33 **CHALLENGE** Consider a binomial distribution in which $n = 50$ and $\sigma = 1.54$. What is the mean of the distribution? (*Hint*: p is closer to 0 than 1.)

34. **WRITING IN MATH** Describe another way you could find the probability that at least three of the teenagers exercise regularly or $P(X \geq 3)$ from Example 7. Give an example of when this method would be faster to use.

Spiral Review

35. ART The prices in dollars of paintings sold at an art auction are shown. (Lesson 11-1)

a. Construct a histogram, and use it to describe the shape of the distribution.

b. Summarize the center and spread of the data using either the mean and standard deviation or the five-number summary. Justify your choice.

Art Prices ($)

1800	600	750	600	600	1800
1350	450	300	1200	750	600
750	450	2700	600	750	300
750	2300	600	450	2100	1200

Use the fifth partial sum of the exponential series to approximate each value to three decimal places. (Lesson 10-6)

36. $e^{0.2}$

37. $e^{-0.4}$

38. $e^{-0.75}$

Find the indicated geometric means for each pair of nonconsecutive terms. (Lesson 10-3)

39. 8 and 312.5; 3 means

40. $\frac{2}{9}$ and 54; 4 means

41. $\frac{3}{4}$ and $\frac{24}{3125}$; 4 means

Find the next four terms of each sequence. (Lesson 10-1)

42. $a_1 = -12, a_n = a_{n-1} + 3, n \geq 2$

43. $a_1 = 19, a_n = a_{n-1} - 13, n \geq 2$

44. $a_1 = 81, a_n = a_{n-1} - 72, n \geq 2$

Find the dot product of u and v. Then determine if u and v are orthogonal. (Lesson 8-5)

45. $\mathbf{u} = \langle 2, 9, -2 \rangle, \mathbf{v} = \langle -4, 7, 6 \rangle$

46. $\mathbf{u} = 3\mathbf{i} - 5\mathbf{j} + 6\mathbf{k}$ and $\mathbf{v} = -7\mathbf{i} + 8\mathbf{j} + 9\mathbf{k}$

47. $\mathbf{u} = \langle 8, -2, -2 \rangle, \mathbf{v} = \langle -6, 6, -10 \rangle$

Graph the hyperbola given by each equation. (Lesson 7-3)

48. $\frac{(y+6)^2}{36} - \frac{(x-1)^2}{24} = 1$

49. $\frac{(y+5)^2}{49} - \frac{(x-6)^2}{20} = 1$

50. $\frac{(y+3)^2}{9} - \frac{(x+5)^2}{4} = 1$

Find *AB* and *BA*, if possible. (Lesson 6-2)

51. $A = [2, -1], B = \begin{bmatrix} 5 \\ 4 \end{bmatrix}$

52. $A = \begin{bmatrix} 3 & -2 \\ 5 & 1 \end{bmatrix}, B = \begin{bmatrix} 4 & 1 \\ 2 & 7 \end{bmatrix}$

53. $A = \begin{bmatrix} 4 & -1 \\ 6 & 1 \\ 5 & -8 \end{bmatrix}, B = \begin{bmatrix} 2 & 9 & -3 \\ 4 & -1 & 0 \end{bmatrix}$

Skills Review for Standardized Tests

54. REVIEW Find the sum of 16 + 8 + 4 + …

A 28

B 32

C 48

D 64

55. In a recent poll, 48% of Americans said that they shopped online for at least one holiday gift. If a random sample of 10 Americans is selected, what is the probability that at least 7 shopped online for a gift?

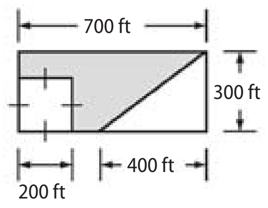
F 3.4%

G 4.8%

H 10.0%

J 14.1%

56. SAT/ACT Find the area of the shaded region.



A 90,000 ft^2 **C** 130,000 ft^2 **E** 210,000 ft^2

B 110,000 ft^2 **D** 150,000 ft^2

57. REVIEW Which of the following distributions best describes the data?

{14, 15, 11, 13, 13, 14, 15, 14, 12, 13, 14, 15}

F positively skewed **H** normal

G negatively skewed **J** binomial

LESSON 11-3 The Normal Distribution

Then

- You analyzed probability distributions for discrete random variables. (Lesson 11-2)

Now

1. Find area under normal distribution curves.
2. Find probabilities for normal distributions, and find data values given probabilities.

Why?

- In a recent year, approximately 107 million Americans 20 years and older had a total blood cholesterol level of 200 milligrams per deciliter or higher. Physicians use variables of this type to compare patients' cholesterol levels to *normal* cholesterol ranges. In this lesson, you will determine the probability of a randomly selected person having a specific cholesterol level.

NewVocabulary
normal distribution
empirical rule
z-value
standard normal distribution

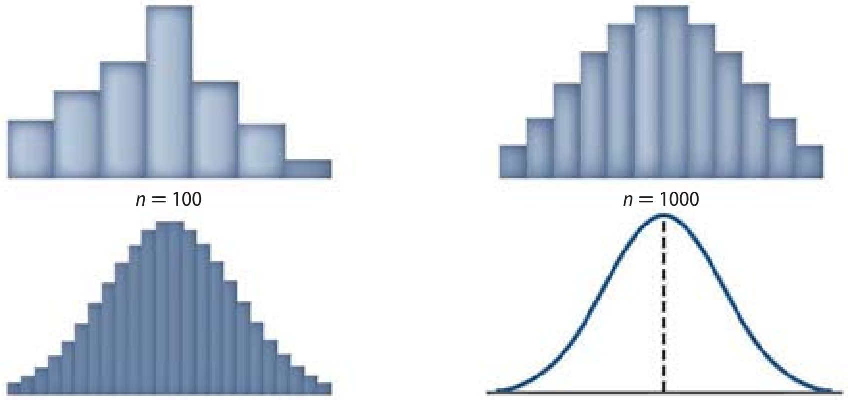
1 The Normal Distribution

The probability distribution for a continuous variable is called a *continuous probability distribution*. The most widely used continuous probability distribution is called the **normal distribution**. The characteristics of the normal distribution are as follows.

KeyConcept Characteristics of the Normal Distribution

- The graph of the curve is bell-shaped and symmetric with respect to the mean.
- The mean, median, and mode are equal and located at the center.
- The curve is continuous.
- The curve approaches, but never touches, the x-axis.
- The total area under the curve is equal to 1 or 100%.

Consider a continuous probability distribution of times for a 400-meter run in a random sample of 100 athletes. By increasing sample size and decreasing class width, the distribution becomes more and more symmetrical. If it were possible to sample the entire population, the distribution would approach the normal distribution, as shown.



For every normally distributed random variable, the shape and position of the normal distribution curve are dependent on the mean and standard deviation. For example, in Figure 11.3.1, you can see that a larger standard deviation results in a flatter curve. A change in the mean, as shown in Figure 11.3.2, results in a horizontal translation of the curve.

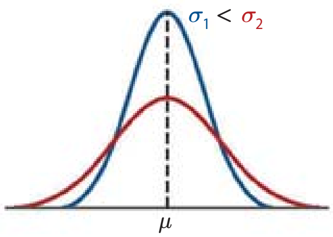


Figure 11.3.1

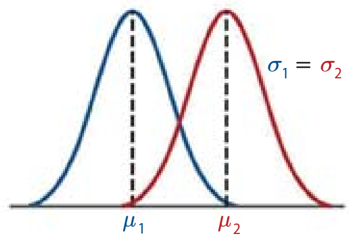


Figure 11.3.2

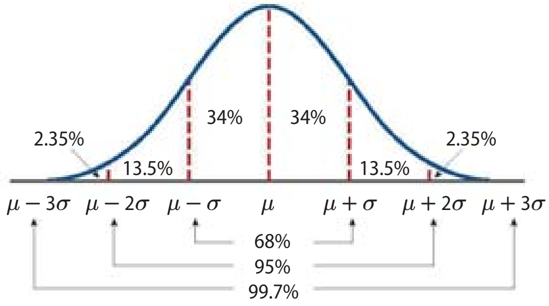
StudyTip

Empirical Rule The empirical rule is also known as the *68–95–99.7 rule.*

The area under the normal distribution curve between two data values represents the percent of data values that fall within that interval. The **empirical rule** can be used to describe areas under the normal curve over intervals that are one, two , or three standard deviations from the mean.

KeyConcept The Empirical Rule

In a normal distribution with mean μ and standard deviation σ:



- approximately 68% of the data values fall between $\mu - \sigma$ and $\mu + \sigma$.
- approximately 95% of the data values fall between $\mu - 2\sigma$ and $\mu + 2\sigma$.
- approximately 99.7% of the data values fall between $\mu - 3\sigma$ and $\mu + 3\sigma$.

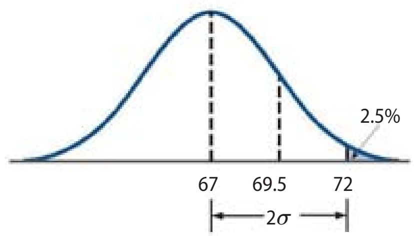
You can solve problems involving approximately normal distributions using the empirical rule.

Example 1 Use the Empirical Rule

HEIGHT The heights of the 880 students at East High School are normally distributed with a mean of 67 inches and a standard deviation of 2.5 inches.

a. Approximately how many students are more than 72 inches tall?

To determine the number of students that are more than 72 inches tall, find the corresponding area under the curve.



In the figure shown, you can see that 72 is 2σ from the mean. Because 95% of the data values fall within two standard deviations from the mean, each tail represents 2.5% of the data. The area to the right of 72 is 2.5% of 880 or 22.

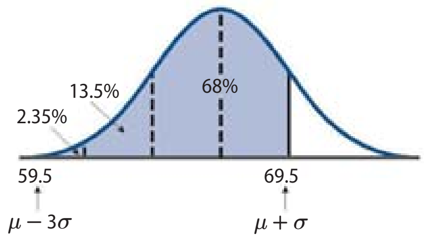
Thus, about 22 students are more than 72 inches tall.

b. What percent of the students are between 59.5 and 69.5 inches tall?

StudyTip

Everything Under the Curve Notice that in Example 1a, we used 2.5%, while in Example 1b, we used 2.35%. When you are asked for *greater than* or *less than*, you need to include everything under that side of the graph.

The percent of students between 59.5 and 69.5 inches tall is represented by the shaded area in the figure at the right, which is between $\mu - 3\sigma$ and $\mu + \sigma$. The total area under the curve between 59.5 and 69.5 is equal to the sum of the areas of each region.



2.35% + 13.5% + 68% = 83.85%

Therefore, about 84% of the students are between 59.5 and 69.5 inches tall.

GuidedPractice

1. **MANUFACTURING** A machine used to fill water bottles dispenses slightly different amounts into each bottle. Suppose the volume of water in 120 bottles is normally distributed with a mean of 1.1 liters and a standard deviation of 0.02 liter.
   - **A.** Approximately how many bottles of water are filled with less than 1.06 liters?
   - **B.** What percent of the bottles have between 1.08 and 1.14 liters?

While the empirical rule can be used to analyze a normal distribution, it is only useful when evaluating specific values, such as $\mu + \sigma$. A normally distributed variable can be transformed into a standard value or z-value, which can be used to analyze any range of values in the normal distribution. This transformation is known as *standardizing*. The **z-value**, also known as the z-score and z test statistic, represents the number of standard deviations that a given data value is from the mean.

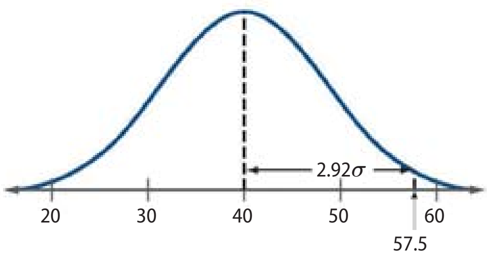
StudyTip

Positive and Negative z-Values If a data value is less than the mean, the corresponding z-value will be negative. Alternately, a data value that is greater than the mean will have a positive z-value.

KeyConcept Formula for z-Values

The z-value for a data value in a set of data is given by $z = \frac{X - \mu}{\sigma}$, where X is the data value, μ is the mean, and σ is the standard deviation.

You can use z-values to determine the position of *any* data value within a set of data. For example, consider a distribution with $\mu = 40$ and $\sigma = 6$. A data value of 57.5 is located about 2.92 standard deviations away from the mean, as shown. Therefore, in this distribution, $X = 57.5$ correlates to a z-value of 2.92.



StudyTip

Relative Position Like percentiles, z-values can be used to compare the relative positions of two values in two different sets of data.

Example 2 Find z-Values

Find each of the following.

a. z if $X = 24$, $\mu = 29$, and $\sigma = 4.2$

$$z = \frac{X - \mu}{\sigma} \quad \text{Formula for } z\text{-values}$$

$$= \frac{24 - 29}{4.2} \quad X = 24,\ \mu = 29,\ \text{and } \sigma = 4.2$$

$$\approx -1.19 \quad \text{Simplify.}$$

The z-value that corresponds to $X = 24$ is -1.19. Therefore, 24 is 1.19 standard deviations less than the mean in the distribution.

b. X if $z = -1.73$, $\mu = 48$, and $\sigma = 2.3$

$$z = \frac{X - \mu}{\sigma} \quad \text{Formula for } z\text{-values}$$

$$-1.73 = \frac{X - 48}{2.3} \quad \mu = 48,\ \sigma = 2.3,\ \text{and } z = -1.73$$

$$-3.979 = X - 48 \quad \text{Multiply each side by 2.3.}$$

$$44.021 = X \quad \text{Add 48 to each side.}$$

A z-value of -1.73 corresponds to a data value of approximately 44 in the distribution.

GuidedPractice

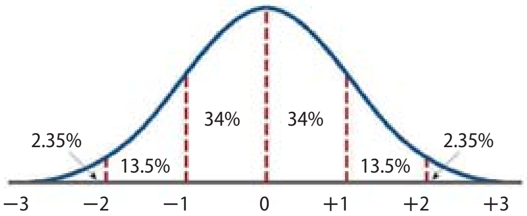
2A. z if $X = 32$, $\mu = 28$, and $\sigma = 1.7$

2B. X if $z = 2.15$, $\mu = 39$, and $\sigma = 0.4$

Every normally distributed random variable has a unique mean and standard deviation, which affect the position and shape of the curve. As a result, there are infinitely many normal probability distributions. Fortunately, they can all be related to one distribution known as the *standard* normal distribution. The **standard normal distribution** is a normal distribution of z-values with a mean of 0 and a standard deviation of 1.

The characteristics of the standard normal distribution are summarized below.

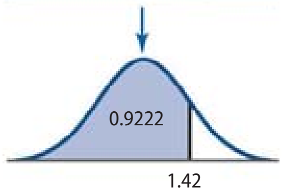
KeyConcept Characteristics of the Standard Normal Distribution



- The total area under the curve is equal to 1 or 100%.
- Almost all of the area is between $z = -3$ and $z = 3$.
- The distribution is symmetric.
- The mean is 0, and the standard deviation is 1.
- The curve approaches, but never touches, the x-axis.

You can solve normal distribution problems by finding the z-value that corresponds to a given X-value, and then finding the approximate area under the standard normal curve. The corresponding area can be found by using a table of z-values that shows the area *to the left* of a given z-value. For example, the area under the curve to the left of a z-value of 1.42 is 0.9222, as shown.

z	0.00	0.01	0.02
0.0	.5000	.5040	.5080
•	•	•	•
•	•	•	•
1.4	.9192	.9207	.9222



You can also find the area under the curve that corresponds to any z-value with a graphing calculator. This method will be used for the remainder of this chapter.

Example 3 Use the Standard Normal Distribution

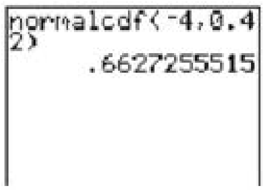
COMMUNICATION The average number of phone calls received by a customer service representative each day during a 30-day month was 105 with a standard deviation of 12. Find the number of days with fewer than 110 phone calls. Assume that the number of calls is normally distributed.

$$z = \frac{X - \mu}{\sigma} \quad \text{Formula for } z\text{-values}$$

$$= \frac{110 - 105}{12} \text{ or about } 0.42 \quad X = 110,\ \mu = 105, \text{ and } \sigma = 12$$

Although the standard normal distribution extends to negative and positive infinity, when you are finding the area less than or greater than a given value, you can use a lower value of −4 and an upper value of 4.

In this case, enter a lower z-value of −4 and an upper z-value of 0.42. The resulting area is 0.66. Since there were 30 days in the month, there were fewer than 110 calls on 30 • 0.66 or 19.8 days.



Therefore, there were approximately 20 days with fewer than 110 calls.

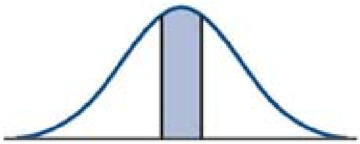
TechnologyTip

Area Under the Normal Curve You can use a graphing calculator to find the area under a standard normal curve that corresponds to any pair of z-values by selecting [2nd] [DISTR] and normalcdf (*lower z value, upper z value*).

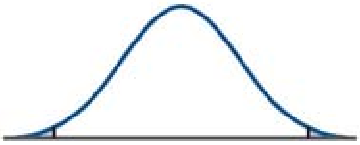
GuidedPractice

3. **BASKETBALL** The average number of points that a basketball team scored during a single season was 63 with a standard deviation of 18. If there were 15 games during the season, find the percentage of games in which the team scored more than 70 points. Assume that the number of points is normally distributed.

In Example 3, you found the area under the normal curve that corresponds to a z-value. You can also find z-values that correspond to specific areas. For example, you can find the z-value that corresponds to a cumulative area of 1%, 20%, or 99%. You can also find intervals of z-values that contain or are between a certain percentage of data.



middle 20%



outside 1%

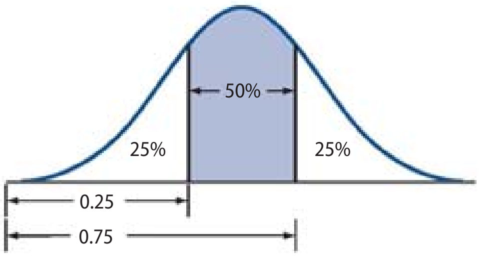
Example 4 Find z-Values Corresponding to a Given Area

> **StudyTip**
>
> **Symmetry** The normal distribution is symmetrical, so when you are asked for the middle or outside set of data, the z-values will be opposites.

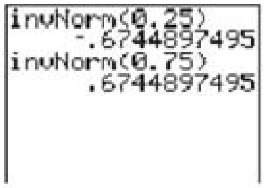
Find the interval of z-values associated with each area.

a. middle 50% of the data

The middle 50% of the data corresponds to the data between 25% and 75% of the distribution, or 0.25 and 0.75, as shown.



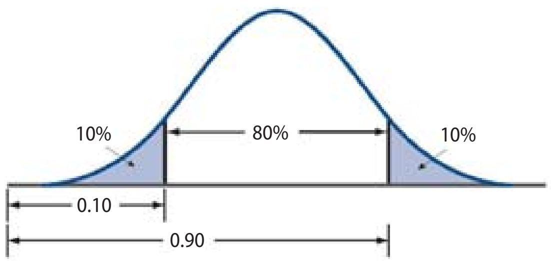
To find the z-scores that correspond to 0.25 and 0.75, select 2nd [DISTR] to display the DISTR menu on a graphing calculator. Select invNorm(and enter 0.25. Repeat to find the value corresponding to 0.75. As shown at the right, the z-value corresponding to 0.25 is -0.67 and the z-value corresponding to 0.75 is 0.67.



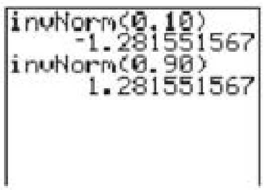
Therefore, the interval that represents the middle 50% of the data is $-0.67 < z < 0.67$.

b. the outside 20% of the data

The outside 20% of the data represents the top 10% and the bottom 10% of the distribution or 0.1 and 0.9, as shown.



To find the z-value corresponding to 0.10, enter 0.10 into a graphing calculator under invNorm(and repeat for 0.90. As shown, the z-value corresponding to 0.10 is -1.28 and the z-value corresponding to 0.90 is 1.28.



Therefore, the interval that represents the outside 20% of the data is $-1.28 > z$ or $z > 1.28$.

GuidedPractice

4A. the middle 25% of the data

4B. the outside 60% of the data

Study Tip

Percentage, Proportion, Probability, and Area When a problem asks for a percentage, proportion, or probability, it is asking for the same value—the corresponding area under the normal curve.

2 Probability and the Normal Distribution

You have seen how the area under the normal curve corresponds to the proportion of data values in an interval. The area also corresponds to the probability of data values falling within a given interval. If a z-value is chosen randomly, the probability of choosing a value between 0 and 1 would be equivalent to the area under the curve between 0 and 1.00, which is 0.3413. Therefore, the probability of choosing a value between 0 and 1 would be approximately 34%.

Example 5 Find Probabilities

METEOROLOGY The temperatures for one month for a city in California are normally distributed with $\mu = 81°$ and $\sigma = 6°$. Find each probability, and use a graphing calculator to sketch the corresponding area under the curve.

a. $P(70° < X < 90°)$

The question is asking for the percentage of temperatures that were between 70° and 90°. First, find the corresponding z-values for $X = 70$ and $X = 90$.

$z = \frac{X - \mu}{\sigma}$ — Formula for z-values

$= \frac{70 - 81}{6}$ — $X = 70$, $\mu = 81$, and $\sigma = 6$

≈ -1.83 — Simplify.

Use 90 to find the other z-value.

$z = \frac{X - \mu}{\sigma}$ — Formula for z-values

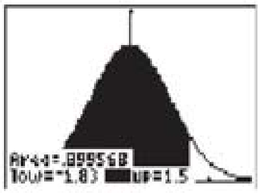
$= \frac{90 - 81}{6}$ — $X = 90$, $\mu = 81$, and $\sigma = 6$

≈ 1.5 — Simplify.

Study Tip

Continuity Factors In a continuous distribution, there is no difference between $P(x \geq c)$ and $P(x > c)$ because the probability that x is equal to c is zero.

You can use a graphing calculator to display the area that corresponds to any z-value by selecting [2nd] [DISTR]. Then, under the DRAW menu, select ShadeNorm (*lower z value, upper z value*). The area between $z = -1.83$ and $z = 1.5$ is 0.899568, as shown.



[−4, 4] scl: 1 by [0, 0.5] scl: 0.125

Therefore, approximately 90% of the temperatures were between 70 and 90.

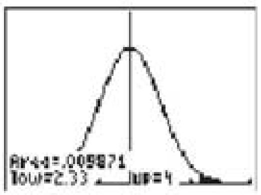
b. $P(X \geq 95°)$

$z = \frac{X - \mu}{\sigma}$ — Formula for z-values

$= \frac{95 - 81}{6}$ — $X = 95$, $\mu = 81$, and $\sigma = 6$

≈ 2.33 — Simplify.

Using a graphing calculator, you can find the area between $z = 2.33$ and $z = 4$ to be 0.0099.



[−4, 4] scl: 1 by [0, 0.5] scl: 0.125

Therefore, the probability that a randomly selected temperature is at least 95° is about 0.1%.

Guided Practice

5. **TESTING** The scores on a standardized test are normally distributed with $\mu = 72$ and $\sigma = 11$. Find each probability and use a graphing calculator to sketch the corresponding area under the curve.

A. $P(X < 89)$ **B.** $P(65 < X < 85)$

You can find specific intervals of data for given probabilities or percentages by using the standard normal distribution.

Real-WorldLink

In a recent year, the average national SAT scores were 502 in Critical Reading, 515 in Math, and 494 in Writing. The average national ACT score in that same year was 21.1.

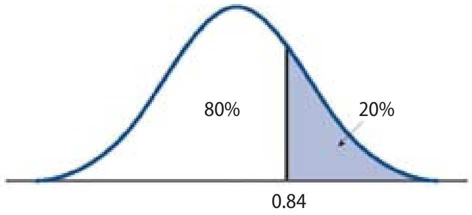
Source: *USA TODAY*

Real-World Example 6 Find Intervals of Data

COLLEGE The scores for the entrance exam for a college's mathematics department is normally distributed with $\mu = 65$ and $\sigma = 8$.

a. If Ramona wants to be in the top 20%, what score must she get?

To find the top 20% of the exam scores, you must find the exam score X that separates the upper 20% of the area under the normal curve, as shown. The top 20% correlates with $1 - 0.2$ or 0.8. Using a graphing calculator, you can find the corresponding z-value to be 0.84.



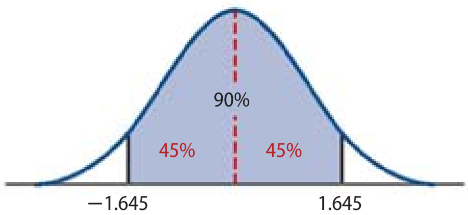
Now, use the formula for the z-value for a population to find the corresponding exam score.

$z = \dfrac{X - \mu}{\sigma}$	Formula for z-values
$0.84 = \dfrac{X - 65}{8}$	$\mu = 65$, $\sigma = 8$, and $z = 0.84$
$6.72 = X - 65$	Multiply each side by 8.
$71.72 = X$	Add 65 to each side.

Ramona needs a score of at least 72 to be in the top 20%.

b. Ramona expects to earn a grade in the middle 90% of the distribution. What range of scores fall in this category?

The middle 90% of the exam scores represents 45% on each side of the mean and therefore corresponds to the interval of area from 0.05 to 0.95. Using a graphing calculator, the z-values that correspond to 0.05 and 0.95 are -1.645 and 1.645, respectively.



Use the z-values to find each value of X.

$z = \dfrac{X - \mu}{\sigma}$	Formula for z-values	$z = \dfrac{X - \mu}{\sigma}$
$-1.645 = \dfrac{X - 65}{8}$	$\mu = 65$ and $\sigma = 8$	$1.645 = \dfrac{X - 65}{8}$
$-13.16 = X - 65$	Multiply.	$13.16 = X - 65$
$51.84 = X$	Simplify.	$78.16 = X$

Therefore, Ramona expects to score between 52 and 78.

GuidedPractice

6. RESEARCH As part of a medical study, a researcher selects a study group with a mean weight of 190 pounds and a standard deviation of 12 pounds. Assume that the weights are normally distributed.

A. If the study will mainly focus on participants whose weights are in the middle 80% of the data set, what range of weights will this include?

B. If participants whose weights fall in the outside 5% of the distribution are contacted 2 weeks after the study, people in what weight range will be contacted?

Exercises

= Step-by-Step Solutions begin on page R29.

1. **NOISE POLLUTION** As part of a noise pollution study, researchers measured the sound level in decibels of a busy city street for 30 days. According to the study, the average noise was 82 decibels with a standard deviation of 6 decibels. Assume that the data are normally distributed. (Example 1)
   a. If a normal conversation is held at about 64 decibels, determine the number of hours during the study that the noise level was this low.
   b. Determine the percent of the study during which the noise was between 76 decibels and 88 decibels.

2. **GAS MILEAGE** Dion commutes 290 miles each week for work. His car averages 29.6 miles per gallon with a standard deviation of 5.4 miles per gallon. Assume that the data are normally distributed. (Example 1)
   a. Approximate the number of miles that Dion's car gets a gas mileage of 35 miles per gallon or better.
   b. For what percentage of Dion's commute does his car have a gas mileage between 24.2 miles per gallon and 40.4 miles per gallon?

Find each of the following. (Example 2)

3. z if $X = 19$, $\mu = 22$, and $\sigma = 2.6$
4. X if $z = 2.3$, $\mu = 64$, and $\sigma = 1.3$
5. z if $X = 52$, $\mu = 43$, and $\sigma = 3.7$
6. X if $z = 2.5$, $\mu = 27$, and $\sigma = 0.4$
7. z if $X = 32$, $\mu = 38$, and $\sigma = 2.8$
8. X if $z = 1.7$, $\mu = 49$, and $\sigma = 4.1$

9. **ICHTHYOLOGY** As part of a science project, José studied the growth rate of 797 green gold catfish and found the following information. Assume that the data are normally distributed. (Example 3)

The green gold catfish reaches its maximum length within its first 3 months of life.

- Average length at birth 4.69 millimeters
- Standard deviation 0.258 millimeters

   a. Determine the number of fish with a length less than 4.5 millimeters at birth.
   b. Determine the number of fish with a length greater than 5 millimeters at birth.

10. **ROLLER COASTER** The average wait in line for the 16,000 daily passengers of a roller coaster is 72 minutes with a standard deviation of 15 minutes. Assume that the data are normally distributed. (Example 3)
   a. Determine the number of passengers who wait less than 60 minutes to ride the roller coaster.
   b. Determine the number of passengers who wait more than 90 minutes to ride the roller coaster.

Find the interval of z-values associated with each area. (Example 4).

11. middle 30%
12. outside 15%
13. outside 40%
14. middle 10%
15. outside 25%
16. middle 84%

17. **BATTERY** The life of a certain brand of AA battery is normally distributed with $\mu = 8$ hours and $\sigma = 1.5$ hours. Find each probability. (Example 5)
   a. The battery will last less than 6 hours.
   b. The battery will last more than 12 hours.
   c. The battery will last between 8 and 9 hours.

18. **HEALTH** The average blood cholesterol level in adult Americans is 203 mg/dL (milligrams per deciliter) with a standard deviation of 38.8 mg/dL. Find each probability. Assume that the data are normally distributed. (Example 5)
   a. a blood cholesterol level below 160 mg/dL, which is considered low and can lead to a higher risk of stroke
   b. a blood cholesterol level above 240 mg/dL, which is considered high and can lead to higher risk of heart disease
   c. a blood cholesterol level between 180 and 200 mg/dL, which is considered normal

19. **SNOWFALL** The average annual snowfall in centimeters for the U.S. and Canada region from 45°N to 55°N is normally distributed with $\mu = 260$ and $\sigma = 27$. (Example 6)
   a. Determine the minimum amount of snowfall occurring in the top 15% of the distribution.
   b. Determine the maximum amount of snowfall occurring in the bottom 30%.
   c. What range of snowfall occurs in the middle 60%?

20. **TRAFFIC SPEED** The speed in miles per hour of traffic on North Street is normally distributed with $\mu = 37.5$ and $\sigma = 5.5$. (Example 6)
   a. Determine the maximum speed of the slowest 10% of cars driving on North Street.
   b. Determine the minimum speed of the fastest 5% of cars driving on North Street.
   c. At what range of speed do the middle 25% of cars on North Street drive?

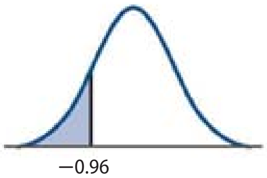
21. **TESTS** Miki took the ACT and SAT and earned the math scores shown. Which of the scores has a higher relative position? Explain your reasoning.

Test	Miki's Score	National Average	Standard Deviation
ACT	27	21	4.7
SAT	620	508	111

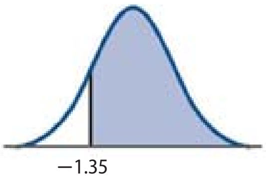
22. EXAMS Bena scored 76 on a physics test that had a mean of 72 and a standard deviation of 10. She also scored 81 on a sociology test that had a mean of 78 and a standard deviation of 9. Compare her relative scores on each test. Assume that the data are normally distributed.

Find the area that corresponds to each shaded region.

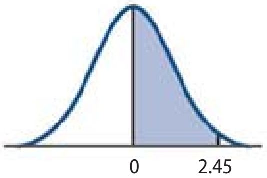
23.



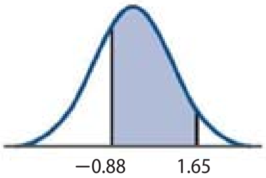
24.



25.



26.



27. FRACTILES Recall from Lesson 11-1 that quartiles, percentiles, and deciles are three types of fractiles, which divide an ordered set of data into equal groups. Find the z-values that correspond to each of the following fractiles.

a. D_{20}, D_{40}, and D_{80}

b. Q_1, Q_2, and Q_3

c. P_{10}, P_{40}, and P_{90}

28. METEOROLOGY The humidity observed in the morning during the same day in Chicago, Orlando, and Phoenix is shown. Assume that the data are normally distributed.

City	Humidity	Average Humidity	Standard Deviation
Chicago	85%	82%	12%
Orlando	94%	91%	15%
Phoenix	46%	43%	10%

a. Which city has the highest humidity? the lowest humidity? Explain your reasoning.

b. How would a fourth city compare that has a humidity of 81% and an average humidity of 78% with a standard deviation of 8%?

29. JOBS The salaries of employees in the sales department of an advertising agency are normally distributed with a standard deviation of \$8000. During the holiday season, employees who earn less than \$35,000 receive a gift basket.

a. Suppose 10% of the employees receive a gift basket. What was the mean salary of the sales department?

b. Suppose employees who make \$10,000 greater than the mean salary receive an incentive bonus. If 200 employees work in the sales department, how many employees will receive a bonus?

30. MULTIPLE REPRESENTATIONS In this problem, you will investigate the shape of a normal distribution. Consider a population of 4, 10, 6, 8.

a. GRAPHICAL Construct a bar graph, and use it to describe the shape of the distribution. Then find the mean and standard deviation of the data set.

b. GRAPHICAL Select eight random samples of size 2, with replacement, from the data set. Construct a bar graph, and use it to describe the shape of the distribution. Find the mean and standard deviation of the sample means.

c. TABULAR The table includes all samples of size 2 that can be taken, with replacement, from the data set. Find the mean of each sample and the mean and standard deviation of all of the sample means.

Sample	Mean	Sample	Mean
4, 4		8, 4	
4, 6		8, 6	
4, 8		8, 8	
4, 10		8, 10	
6, 4		10, 4	
6, 6		10, 6	
6, 8		10, 8	
6, 10		10, 10	

d. GRAPHICAL Construct a bar graph of the sample means from part **c** and use it to describe the shape of the distribution. What happens to the shape of a distribution of data as the sample size increases?

e. ANALYTICAL Divide the standard deviation of the population that you found in part **a** by the square root of the sample size. What do you think happens to the mean and standard deviation of a distribution of data as the sample size increases?

H.O.T. Problems Use Higher-Order Thinking Skills

31. ERROR ANALYSIS Chad and Lucy are finding the z interval associated with the outside 35% of a distribution of data. Chad thinks it is the interval $z < -0.39$ or $z > 0.39$, while Lucy thinks it is the interval $z < -0.93$ or $z > 0.93$. Is either of them correct? Explain your reasoning.

32. REASONING In real-life applications, z-values usually fall between -3 and $+3$ in the standard normal distribution. Why do you think this is the case? Explain your reasoning.

33. CHALLENGE Find two z-values, one positive and one negative, so that the combined area of the two equivalent tails is equal to each of the following.

a. 1% **b.** 5% **c.** 10%

34. REASONING Continuous variables *sometimes*, *always*, or *never* have normal distributions. Explain your reasoning.

35. WRITING IN MATH Compare and contrast the characteristics of a normal distribution and the standard normal distribution.

Spiral Review

36. BASEBALL The number of hits by each Wildcats player during a doubleheader is shown in the frequency distribution. (Lesson 11-2)

Hits, X	Frequency
0	3
1	1
2	8
3	2
4	3

a. Construct and graph a probability distribution for the random variable X.

b. Find and interpret the mean in the context of the situation.

c. Find the variance and standard deviation.

37. FOOTBALL The number of penalties a professional football team received for each game during two recent seasons is shown. Construct side-by-side box plots of the data sets. Then use this display to compare the distributions. (Lesson 11-1)

Season 1				Season 2			
8	11	6	13	9	1	3	5
9	18	16	11	8	3	6	4
15	14	14	9	10	6	3	1
8	5	10	5	5	5	3	2

Find the sum of each arithmetic series. (Lesson 10-2)

38. S_{51} of $-92 + (-88) + (-84) + \ldots$

39. 24th partial sum of $-13 + 2 + 17 + \ldots$

40. S_{46} of $295 + 281 + 267 + \ldots$

Find rectangular coordinates for each point with the given polar coordinates. (Lesson 9-3)

41. $\left(\frac{1}{4}, \frac{\pi}{2}\right)$

42. $\left(3, \frac{\pi}{3}\right)$

43. $(-2, \pi)$

Given v and u • v, find u. There may be more than one answer. (Lesson 8-5)

44. $\mathbf{v} = \langle -4, 2, -7 \rangle$, $\mathbf{u} \cdot \mathbf{v} = 17$

45. $\mathbf{v} = \langle 2, 8, 5 \rangle$, $\mathbf{u} \cdot \mathbf{v} = -6$

46. $\mathbf{v} = \left\langle \frac{2}{3}, -3, \frac{1}{3} \right\rangle$, $\mathbf{u} \cdot \mathbf{v} = 10$

Find the direction angle of each vector. (Lesson 8-2)

47. $6\mathbf{i} + 3\mathbf{j}$

48. $-3\mathbf{i} + 4\mathbf{j}$

49. $2\mathbf{i} - 8\mathbf{j}$

Write an equation of an ellipse with each set of characteristics. (Lesson 7-2)

50. vertices $(-3, 11)$, $(-3, -9)$; foci $(-3, 7)$, $(-3, -5)$

51. co-vertices $(-1, -6)$, $(-3, -6)$; length of major axis equals 10

52. vertices $(-4, 2)$, $(8, 2)$; length of minor axis equals 8

Skills Review for Standardized Tests

53. SAT/ACT If X is the sum of the first 1000 positive even integers and Y is the sum of the first 500 positive odd integers, about what percent greater is X than Y?

A 100% **C** 300% **E** 500%

B 200% **D** 400%

54. REVIEW In a recent year, the mean and standard deviation for scores on the ACT was 21.0 and 4.7. Assume that the scores were normally distributed. What is the approximate probability that a test taker scored higher than 30.2?

F 1% **H** 2%

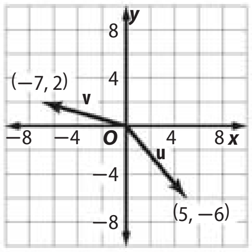
G 1.5% **J** 2.5%

55. The length of each song in a music collection is normally distributed with $\mu = 4.12$ minutes and $\sigma = 0.68$ minutes. Find the probability that a song selected from the collection at random is longer than 5 minutes.

A 10% **C** 39%

B 19% **D** 89%

56. REVIEW Find $\mathbf{u} \cdot \mathbf{v}$.



F −47 **H** −6

G −24 **J** 47

EXTEND

11-3 Graphing Technology Lab
Transforming Skewed Data

Objective

- Use a graphing calculator to transform skewed data into data that resemble a normal distribution.

It is common for biological, medical, and other data to be positively skewed. It can sometimes be helpful to *transform* the original data so that it better resembles a normal distribution. This allows for the data to be spread out as opposed to being bunched at one end of a display.

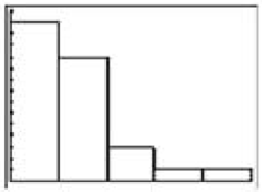
Activity Transform Data Using Natural Logarithms

Use the data to construct a histogram, and describe the shape of the distribution. Then transform the data by calculating the common logarithm of each entry. Graph the new data, and describe the shape of the distribution.

Data									
15	7	2	5	8	17	15	8	3	4
9	18	13	10	9	8	10	23	26	10
7	14	25	7	6	13	35	48	14	6

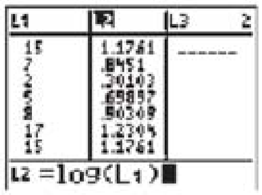
Step 1 Input the data into L1. Construct a histogram for the data using the intervals and scales shown.

The data appear to be positively skewed.



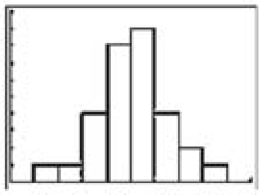
[0, 50] scl: 10 by [0, 15] scl: 1

Step 2 Input the common logarithm for each value into L2. Place the cursor on L2. Press LOG and enter L1. Press ENTER.



Step 3 Construct a histogram for the new data using the intervals and scales shown.

The data appear to have a normal distribution.



[0, 2] scl: 0.2 by [0, 10] scl: 1

Data may also be transformed by calculating the square roots or powers of the entries. When data are transformed, the type of operation performed should always be specified. A transformation will not always result in the new data being normally distributed.

Exercise

Use the data to construct a histogram, and describe the shape of the distribution. Then transform the data by calculating the square root of each entry. Graph the new data, and describe the shape of the distribution. Explain how the transformation affected the summary statistics.

Data									
23	30	36	39	36	24	31	33	42	36
26	32	46	45	27	34	52	41	28	33
43	20	24	34	30	40	29	35	61	35

LESSON 11-4

The Central Limit Theorem

Then	Now	Why?
You used the normal distribution to find probabilities for intervals of data values in distributions. (Lesson 11-3)	1 Use the Central Limit Theorem to find probabilities. 2 Find normal approximations of binomial distributions.	In manufacturing processes, quality control systems are used to determine when a process is outside of upper and lower control limits or "out of control." The mean of the process is controlled; therefore, successive sample means should be normally distributed around the actual mean.

NewVocabulary
- sampling distribution
- standard error of the mean
- sampling error
- continuity correction factor

1 The Central Limit Theorem

Sampling is an important statistical tool in which subgroups of a population are selected so that inferences can be made about the entire population. The means of these subgroups, or sample means, can be compared to the mean of the population by using a sampling distribution. A **sampling distribution** is a distribution of the means of random samples of a certain size that are taken from a population.

Consider a population consisting of 16, 18, 20, 20, 22, and 24, with $\mu = 20$ and $\sigma = 2.582$. Suppose 12 random samples of size 2 are taken, with replacement. The mean $\bar{x}$ of each sample is shown.

Sample	$\bar{x}$	Sample	$\bar{x}$	Sample	$\bar{x}$
20,22	21	20,18	19	22,22	22
22,18	20	16,22	19	18,18	18
20,24	22	24,16	20	20,16	18
20,20	20	20,24	22	24,22	23

The distribution of the means of the 12 random samples, shown in Figure 11.4.1, does not appear to be normal. However, if all 36 samples of size of 2 from the population are found, the distribution of sample means will approach the normal distribution, as shown in Figure 11.4.2.

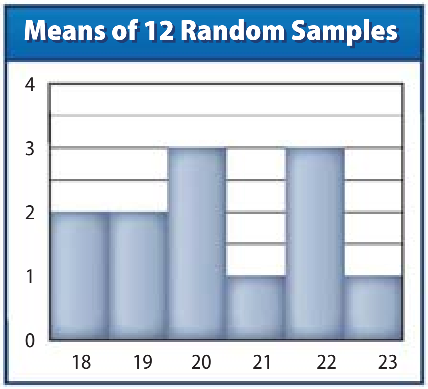


Figure 11.4.1

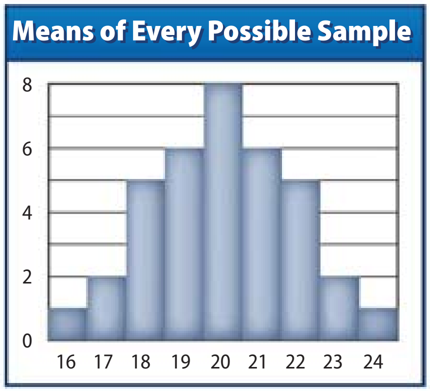


Figure 11.4.2

The mean of the means of every possible sample of size 2 from the population is

$$\mu_{\bar{x}} = \frac{16 + 17 + \ldots + 24}{36} = \frac{720}{36} \text{ or } 20.$$

Notice that this value is equal to the population mean $\mu = 20$. So, when the mean of the means of every possible sample of size 2 are found, $\mu_{\bar{x}} = \mu$. The standard deviation of the sample means $\sigma_{\bar{x}}$ and the standard deviation of the population σ when divided by the square root of the sample of size n are

$$\sigma_{\bar{x}} = \frac{\sqrt{(16-20)^2 + (17-20)^2 + \ldots + (24-20)^2}}{36} \approx 1.826 \quad \text{and} \quad \frac{\sigma}{\sqrt{n}} = \frac{2.582}{\sqrt{2}} \approx 1.826.$$

Since these two values are equal, the standard deviation of the sample means, also known as the **standard error of the mean**, can be found by using the formula $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$.

In general, randomly selected samples will have sample means that differ from the population mean. These differences are caused by **sampling error**, which occurs because the sample is not a complete representation of the population. However, if *all* possible samples of size n are taken from a population with mean μ and a standard deviation σ, the distribution of sample means will have:

- a mean $\mu_{\bar{x}}$ that is equal to μ and
- a standard deviation $\sigma_{\bar{x}}$ that is equal to $\frac{\sigma}{\sqrt{n}}$.

When the sample size n is large, regardless of the shape of the original distribution, the Central Limit Theorem states that the shape of the distribution of the sample means will approach a normal distribution.

StudyTip

Normally Distributed Variables If the original variable is not normally distributed, then n must be greater than 30 in order to use the standard normal distribution to approximate a distribution of sample means.

KeyConcept Central Limit Theorem

As the sampling size n increases:

- the shape of the distribution of the sample means of a population with mean μ and standard deviation σ will approach a normal distribution and
- the distribution will have a mean μ and standard deviation $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$.

The Central Limit Theorem can be used to answer questions about sample means in the same way that the normal distribution was used to answer questions about individual values. In this case, we can use a formula for the z-value of a sample mean.

KeyConcept z-Value of a Sample Mean

The z-value for a sample mean in a population is given by $z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}}$, where $\bar{x}$ is the sample mean, μ is the mean of the population, and $\sigma_{\bar{x}}$ is the standard error.

Example 1 Use the Central Limit Theorem

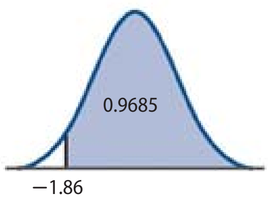
AGE According to a recent study, the average age that an American adult leaves home is 26 years old. Assume that this variable is normally distributed with a standard deviation of 2.4 years. If a random sample of 20 adults is selected, find the probability that the mean age the participants left home is greater than 25 years old.

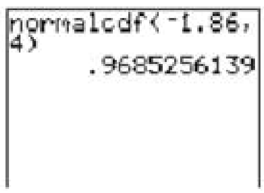
Since the variable is normally distributed, the distribution of the sample means will be approximately normal with $\mu = 26$ and $\sigma_{\bar{x}} = \frac{2.4}{\sqrt{20}}$ or about 0.537. Find the z-value.

$z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}}$ — z-value for a sample mean

$= \frac{25 - 26}{0.537}$ — $\bar{x} = 25$, $\mu = 26$, and $\sigma_{\bar{x}} = 0.537$

≈ -1.86 — Simplify.





The area to the right of a z-value of -1.86 is 0.9685. Therefore, the probability that the mean age of the sample is greater than 25 or $P(\bar{x} > 25)$ is about 96.85%.

GuidedPractice

1. **TORNADOES** The average number of tornadoes in Kansas is 47 per year, with a standard deviation of approximately 14.2 tornadoes. If a random sample of 15 years is selected, find the probability that the mean number of tornadoes is less than 50.

Real-WorldLink

In 1994, a nonprofit organization called The Rechargeable Battery Recycling Corporation was formed to promote the recycling of rechargeable batteries in North America. It provides information for over 50,000 collection locations nationwide where rechargeable batteries can be recycled for free.

Source: Battery University

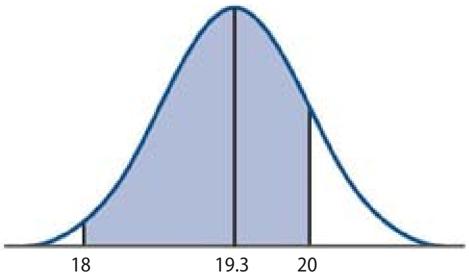
Don Farrall/Photodisc/Getty Images

You can also determine the probability that a sample mean will fall within a given interval of the sampling distribution.

Real-World Example 2 Find the Area Between Two Sample Means

BATTERY LIFE A company that produces rechargeable batteries is designing a battery that will need to be recharged after an average of 19.3 hours of use. Assume that the distribution is normal with a standard deviation of 2.4 hours. If a random sample of 20 batteries is selected, find the probability that the mean life of the batteries before recharging is between 18 and 20 hours.

The area that corresponds to an interval of 18 to 20 hours is shown at the right.



First, find the standard deviation of the sample means.

$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$	Standard deviation of a sample mean
$= \frac{2.4}{\sqrt{20}}$	$\sigma = 2.4$ and $n = 20$
≈ 0.536	Simplify.

Use the z-value formula for a sample mean to find the corresponding z-values for 18 and 20.

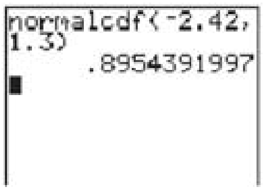
z-value for $\bar{x} = 18$:

$z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}}$	z-value formula for a sample mean
$= \frac{18 - 19.3}{0.536}$	$\bar{x} = 18$, $\mu = 19.3$, and $\sigma_{\bar{x}} = 0.536$
≈ -2.42	Simplify.

z-value for $\bar{x} = 20$:

$z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}}$	z-value formula for a sample mean
$= \frac{20 - 19.3}{0.536}$	$\bar{x} = 20$, $\mu = 19.3$, and $\sigma_{\bar{x}} = 0.536$
≈ 1.30	Simplify.

Using a graphing calculator, select normalcdf(to find the area between $z = -2.42$ and $z = 1.30$.



The area between z-values of -2.42 and 1.30 is 0.8954. Therefore, $P(18 < \mu < 20)$ is 89.54%. So, the probability that the mean life of the batteries is between 18 and 20 hours is 89.54%.

GuidedPractice

2. **DAIRY** The average cost of a gallon of milk in a U.S. city is \$3.49 with a standard deviation of \$0.24. If a random sample of 40 1-gallon containers of milk is selected, find the probability that the mean of the sample will be between \$3.40 and \$3.60.

Example 3 Analyze Individual Values and Sample Means

CLASS SIZE According to a recent study, the average class size in high schools nationwide is 24.7 students per class. Assume that the distribution is normal with a standard deviation of 3.6 students.

a. Find the probability that a randomly selected class will have fewer than 23 students.

The question is asking for an individual value in which $P(x < 23)$. Use the z-value formula for an individual data value to find the corresponding z-value.

$$z = \frac{X - \mu}{\sigma} \quad \text{z-value formula for an individual value}$$

$$= \frac{23 - 24.7}{3.6} \text{ or about } -0.47 \quad X = 23, \mu = 24.7, \text{ and } \sigma = 3.6$$

The area associated with $z < -0.47$, or $P(z < -0.47)$, is 0.3192. Therefore, the probability that a randomly selected class has fewer than 23 students is 31.9%.

b. If a sample of 15 classes is selected, find the probability that the mean of the sample will be fewer than 23 students per class.

This question deals with a sample mean, so use the z-value formula for a sample mean to find the corresponding z-value. First, find the standard error of the mean.

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \quad \text{Standard error of the mean}$$

$$= \frac{3.6}{\sqrt{15}} \text{ or about } 0.93 \quad \sigma = 3.6 \text{ and } n = 15$$

Next, find the z-value using the z-value formula for a sample mean.

$$z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} \quad \text{z-value formula for a sample mean}$$

$$= \frac{23 - 24.7}{0.93} \text{ or about } -1.83 \quad \bar{x} = 23, \mu = 24.7, \text{ and } \sigma_{\bar{x}} = 0.93$$

The area associated with $z < -1.83$, or $P(z < -1.83)$, is 0.0336. Therefore, the probability that a sample of 15 classes will have a mean class size of fewer than 23 students is 3.36%.

StudyTip

z-Value Formulas Notice that the difference between the z-value formula for an individual data value and the z-value formula for a sample mean is that $\bar{x}$ is substituted for X and $\sigma_{\bar{x}}$ is substituted for σ in the formula for an individual value.

GuidedPractice

3. APPLES Consumers in the U.S. eat an average of 19 pounds of apples per year. Assume that the standard deviation is 4 pounds and the distribution is approximately normal.

A. Find the probability that a randomly selected person consumes more than 21 pounds of apples per year.

B. If a sample of 30 people is selected, find the probability that the mean of the sample would be more than 21 pounds of apples per year.

Notice in Figure 11.4.3 that the probability that an individual class has fewer than 23 students is much greater than the probability associated with the mean of a sample being fewer than 23 shown in Figure 11.4.4. This means that as the sample size increases, the distribution becomes narrower and the variability decreases.

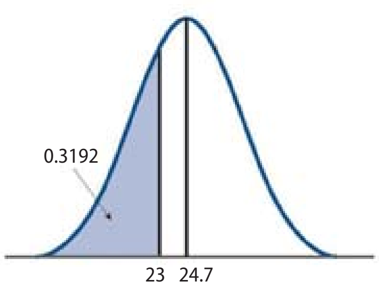


Figure 11.4.3

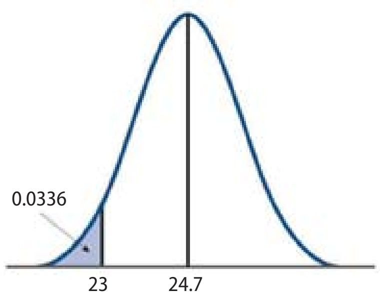


Figure 11.4.4

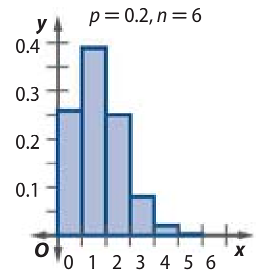
2 The Normal Approximation

According to the Central Limit Theorem, any sampling distribution can approach the normal distribution as n increases. As a result, other distributions such as the binomial distribution can be approximated with the normal distribution. Recall from Lesson 11-2 that the binomial distribution can be determined by using the equation

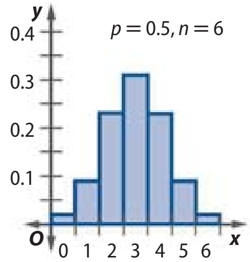
$$P(X) = {}_nC_x\, p^x q^{n-x},$$

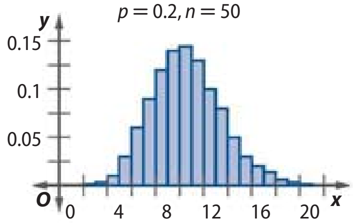
where n is the number of trials, p is the probability of success, and q is the probability of failure.

If the number of trials increases or the probability of success gets close to 0.5, the shape of the binomial distribution begins to resemble the normal distribution. For example, consider the binomial distribution at the right. When $p = 0.2$ and $n = 6$, the distribution is positively skewed.



However, when $p = 0.5$ and $n = 6$ or when $p = 0.2$ and $n = 50$, as shown below, the distribution is approximately normal.





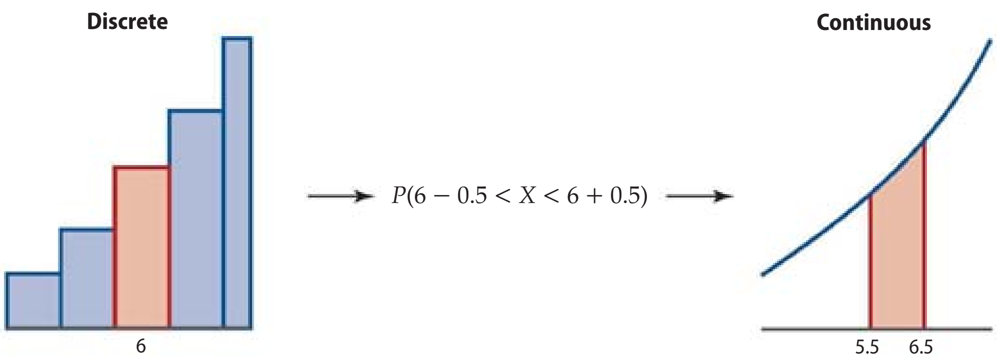
When the probability of success is close to 0 or 1 and the number of trials is relatively small, the normal approximation is not accurate. Therefore, as a rule, the normal approximation is typically used only when $np \geq 5$ and $nq \geq 5$.

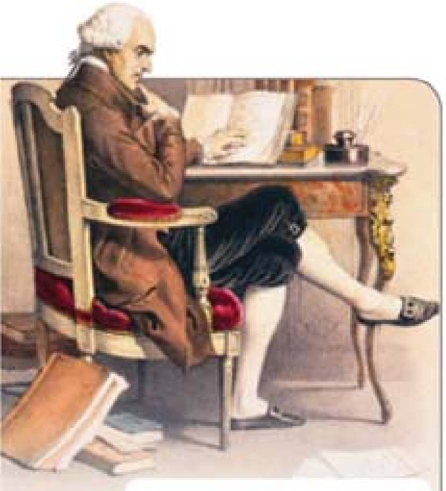
KeyConcept Approximation Rule for Binomial Distributions

Words	The normal distribution can be used to approximate a binomial distribution when $np \geq 5$ and $nq \geq 5$.
Example	If p is 0.4 and n is 5, then $np = 5(0.4)$ or 2. Since $2 < 5$, the normal distribution should not be used to approximate the binomial distribution.

It also is important to remember that the normal distribution should only be used to approximate a binomial distribution if the original variable is normally distributed or $n \geq 30$.

Since binomial distributions are *discrete* and normal distributions are *continuous*, a correction for continuity called the **continuity correction factor** must be used when approximating a binomial distribution. To use the correction factor, 0.5 unit is added to or subtracted from a given discrete boundary. For example, to find $P(X = 6)$ in a discrete distribution, the correction would be to find $P(5.5 < X < 6.5)$ for a continuous distribution, as shown below.





Math HistoryLink

Pierre-Simon Laplace
(1749–1827)
A French mathematician and astronomer, Pierre-Simon Laplace was born in Beaumont-en Auge, France. Laplace first approximated the binomial distribution with the normal distribution in his 1812 work *Théorie Analytique des Probabilités.*

Use the following steps to approximate a binomial distribution with the normal distribution.

StudyTip

Binomial Formulas Recall from Lesson 11-2 that the mean μ and standard deviation σ of a binomial distribution are found using $\mu = np$ and $\sigma = \sqrt{npq}$, respectively.

KeyConcept Normal Approximation of a Binomial Distribution

The procedure for the normal approximation of a binomial distribution is as follows.

Step 1 Find the mean μ and standard deviation σ.

Step 2 Write the problem in probability notation using X.

Step 3 Find the continuity correction factor, and rewrite the problem to show the corresponding area under the normal distribution.

Step 4 Find any corresponding z-values for X.

Step 5 Use a graphing calculator to find the corresponding area.

Example 4 Normal Approximation of a Binomial Distribution

COLLEGE A school newspaper reported that 20% of the current senior class would be attending an out-of-state college. If 35 seniors are selected at random, find the probability that fewer than 5 of the seniors will be attending an out-of-state college.

In this binomial experiment, $n = 35$, $p = 0.2$, and $q = 0.8$.

Step 1 Find the mean μ and standard deviation σ.

$\mu = np$	Mean and standard deviation of a binomial distribution	$\sigma = \sqrt{npq}$
$= 35 \cdot 0.2$	$n = 35$, $p = 0.2$, and $q = 0.8$	$= \sqrt{35 \cdot 0.2 \cdot 0.8}$
$= 7$	Simplify.	≈ 2.37

Since $np = 35(0.2)$ or 7 and $nq = 35(0.8)$ or 28, which are both greater than 5, the normal distribution can be used to approximate the binomial distribution.

Step 2 Write the problem in probability notation using X.

The probability that fewer than 5 of the seniors will be attending an out-of-state college is $P(X < 5)$.

Step 3 Rewrite the problem with the continuity factor included.

Since the question is asking for the probability that *fewer than* 5 will be attending, subtract 0.5 unit from 5.

$P(X < 5) = P(X < 5 - 0.5)$ or $P(X < 4.5)$

WatchOut!

***z*-Value Formula** When approximating the binomial distribution using the normal distribution, remember to use the z-value formula for an individual data value, not the formula for a sample mean.

Step 4 Find the corresponding z-value for X.

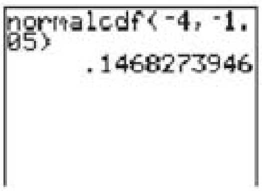
$$z = \frac{X - \mu}{\sigma} \quad \text{z-value formula}$$

$$= \frac{4.5 - 7}{2.37} \quad X = 4.5,\ \mu = 7, \text{ and } \sigma = 2.37$$

$$\approx -1.05 \quad \text{Simplify.}$$

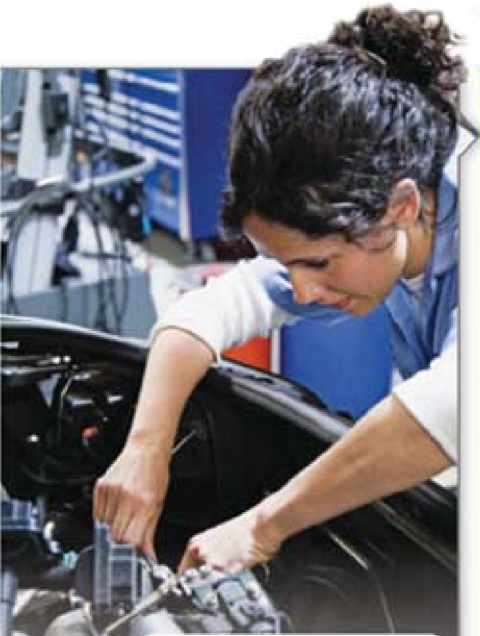
Step 5 Use a graphing calculator to find the area to the left of z.

The approximate area to the left of $z = -1.05$ is 0.147, as shown at the right. Therefore, the probability that fewer than 5 seniors will be attending an out-of-state college in a random sample of 35 seniors is about 14.7%.



GuidedPractice

4. ADVERTISING According to the results of an advertising survey sent to customers selected at random, 65% of the customers had not seen a recent television advertisement. Find the probability that from a sample of 50 customer responses, 15 or more did not see the advertisement.



Real-World Example 5 Normal Approximation of a Binomial Distribution

MANUFACTURING An automaker has discovered a defect in a new model. The defect is expected to affect 30% of the cars that were produced. What is the probability that there are at least 10 and at most 15 cars with the defect in a random sample of 40 cars?

In this binomial experiment, $n = 40$, $p = 0.3$, and $q = 0.7$.

Step 1 Begin by finding the mean μ and standard deviation σ.

$\mu = np$	Mean and standard deviation of a binomial distribution	$\sigma = \sqrt{npq}$
$= 40 \cdot 0.3$	$n = 40$, $p = 0.3$, and $q = 0.7$	$= \sqrt{40 \cdot 0.3 \cdot 0.7}$
$= 12$	Simplify.	≈ 2.9

Since $np = 40(0.3)$ or 12 and $nq = 40(0.7)$ or 28, which are both greater than 5, the normal distribution can be used to approximate the binomial distribution.

Step 2 Write the problem in probability notation: $P(10 \le X \le 15)$.

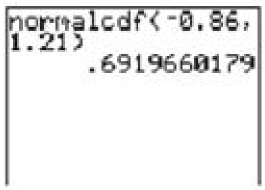
Step 3 Rewrite the problem with the continuity factor included.

$P(10 \le X \le 15) = P(10 - 0.5 < X < 15 + 0.5)$ or $P(9.5 \le X \le 15.5)$

Step 4 Find the corresponding z-values for $X = 9.5$ and $X = 15.5$.

$z = \dfrac{X - \mu}{\sigma}$	z-value formula	$z = \dfrac{X - \mu}{\sigma}$
$= \dfrac{9.5 - 12}{2.9}$	Substitute	$= \dfrac{15.5 - 12}{2.9}$
≈ -0.86	Simplify.	≈ 1.21

Step 5 Use a graphing calculator to find the area between $z = -0.86$ and $z = 1.21$.



The approximate area that corresponds to $-0.86 < z < 1.21$ is 0.692, as shown at the right. Therefore, the probability of there being at least 10 and at most 15 cars with the defect in a random sample of 40 cars is about 69.2%.

GuidedPractice

5. **MANUFACTURING** Suppose a defect in a second model by the same automaker is expected to affect 20% of the cars that were produced. What is the probability that there are at least 8 and at most 10 defects in a random sample of 30 cars?

It may seem difficult to know whether to add or subtract 0.5 unit from a discrete data value to find the continuity correction factor. The table below shows each case.

ConceptSummary Binomial Distribution Correction Factors

When using the normal distribution to approximate a binomial distribution, the following correction factors should be used, where c is a given data value in the binomial distribution.

Binomial	Normal
$P(X = c)$	$P(c - 0.5 < X < c + 0.5)$
$P(X > c)$	$P(X > c + 0.5)$
$P(X \ge c)$	$P(X > c - 0.5)$
$P(X < c)$	$P(X < c - 0.5)$
$P(X \le c)$	$P(X < c + 0.5)$

Real-WorldLink

Product recalls occur when a manufacturer sends out a request to the consumers to return a product after discovering a safety issue. Recalls are costly, but are done to limit the liability of the manufacturer.

Source: National Highway Traffic Safety Administration

WatchOut!

Writing Inequalities When a problem is asking for a probability *between* two values, write the inequality as $P(c_1 < X < c_2)$, not $P(c_1 \le X \le c_2)$. For instance, in Example 5, the probability that there are *between* 10 and 15 defects would be $P(10 < X < 15)$.

Exercises

= Step-by-Step Solutions begin on page R29.

1. **VIDEO GAMES** The average prices for three video games at an online auction site are shown. Assume that the variable is normally distributed. (Examples 1 and 2)

Game	Average Price ($)
Column Craze	35
Dungeon Attack!	45
Space Race	52

   a. For a sample of 35 online prices for Column Craze, find the probability that the mean price is more than $38, if the standard deviation is $9.

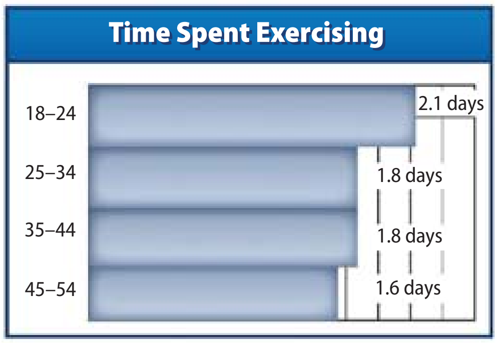
   b. For a random sample of 40 online prices for Space Race, find the probability that the mean price will be between $50 and $55 if the standard deviation is $12.

2. **CHEWING GUM** Americans chew an average of 182 sticks of gum per year. Assume a standard deviation of 13 sticks for each question. Assume that the variable is normally distributed. (Examples 1 and 2)

   a. Find the probability that 50 randomly selected people chew an average of 175 sticks or more per year.

   b. If a random sample of 45 people is selected, find the probability that the mean number of sticks of gum they chew per year is between 180 and 185.

3. **EXERCISE** The average number of days per week that Americans from four different age groups spent exercising during a recent year is shown. Assume that the variable is normally distributed. (Examples 1 and 2)



   a. Find the probability that a random sample of 30 Americans ages 45 to 54 spent more than 1.5 days a week exercising, if the standard deviation is 0.5 day.

   b. Assuming a standard deviation of 1.2 days, in a random sample of 30 Americans ages 18 to 24, find the probability that the average time spent exercising is between 2 to 2.5 days per week.

4. **MEDICINE** The mean recovery time for patients with a certain virus is 4.5 days with a standard deviation of 2 days. Assume that the variable is normally distributed. (Examples 1 and 2)

   a. Find the probability of an average recovery time of less than 4 days for a random sample of 75 people.

   b. In a random sample of 80 people, find the probability that average recovery time is between 4.4 and 4.8 days.

5. **TOURISM** The average number of tourists that visit a national monument every month is 55,000, with a standard deviation of 8000. Assume that the variable is normally distributed. (Example 3)

   a. If a random month is selected, find the probability that there would be fewer than 50,000 visiting tourists.

   b. If a sample of 10 months is selected, find the probability that there would be fewer than 50,000 visiting tourists.

6. **NUTRITION** The average protein content of a certain brand of energy bar is 12 grams with a standard deviation of 2 grams. Assume that the variable is normally distributed. (Example 3)

   a. Find the probability that a randomly selected bar will have more than 10 grams of protein.

   b. In a sample of 15 bars, find the probability that the average protein content will be greater than 10 grams.

7. **WORLD CUP** In a recent year, 33% of Americans said that they were planning to watch the World Cup soccer tournament. What is the probability that in a random sample of 45 people, fewer than 14 people plan to watch the World Cup? Assume that the variable is normally distributed. (Example 4)

8. **MOVIES** According to a national poll, in a recent year, 27% of Americans saw 5 or more movies in theaters. What is the probability that in a random sample of 40 people, between 6 and 11 people saw more than 5 movies in a movie theater that year? Assume that the variable is normally distributed. (Example 5)

9. **LIBRARY** A poll was conducted at a library to approximate the percent of books, CDs, magazines, and movies that were checked out during one month. The results are shown. Assume that the variable is normally distributed. (Examples 4 and 5)

Resources	Percent
books	45
CDs	20
magazines	3
movies	32

   a. What is the probability that of 65 randomly selected resources, exactly 35 were books?

   b. Find the probability that of 85 randomly selected resources, at least 15 and at most 18 were CDs.

10. **DRIVING** A driving instructor has found that 12% of students cancel or forget about lessons. Assume that the variable is normally distributed. (Examples 4 and 5)

   a. If the instructor has 60 students, what is the probability that more than 10 of the students will miss a lesson?

   b. What is the probability that of 80 students, exactly 7 students will miss a lesson?

11. **TESTS** A multiple-choice test consists of 50 questions, with possible answers A, B, C, and D. Find the probability that, with random guessing, the number of correct answers will be each of the following.

a. less than 18 **b.** exactly 12

c. at least 14 **d.** between 10 and 15

Find the minimum sample size needed for each probability so that the normal distribution can be used to approximate the binomial distribution.

12. $p = 0.1$ 13. $p = 0.4$

14. $p = 0.5$ 15. $p = 0.8$

16. **BASKETBALL** The average points per game scored by four different basketball players are shown.

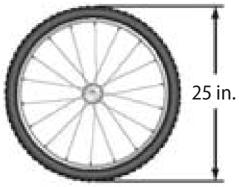
Player	A	B	C	D
Average	8.1	6.3	4.9	10.3

a. Find the mean and standard deviation of the averages.

b. Identify each possible combination of 3 players' averages, and find the mean of each combination.

c. Find the mean of each of the means that you found in part **b.** How does this compare to the mean that you found in part **a**?

17. **BICYCLES** Consider the bicycle rim shown, where $\mu = 25$ inches and $\sigma = 0.125$ inch.



The diameters for 10 random samples of 3 bicycle rims from a company's assembly line are shown.

Sample	Diameter	Sample	Diameter
1	25.2, 24.9, 25	6	24.9, 25.1, 24.8
2	25.1, 25, 24.8	7	25.3, 24.9, 25.1
3	25.3, 24.9, 24.8	8	25.4, 24.8, 25.3
4	24.9, 25.3, 25.2	9	24.8, 24.9, 25.2
5	25, 25.2, 24.7	10	25, 25.3, 24.7

a. Find $\bar{x}$ and s for each sample.

b. Construct a scatter plot with the sample number on the x-axis and the sample means on the y-axis.

c. In this process, the upper control limit is $\bar{x} + \frac{3\sigma}{\sqrt{n}}$ and the lower control limit is $\bar{x} - \frac{3\sigma}{\sqrt{n}}$. If the process is in control, all values should fall within the control limits. Use the graph from part **b** to determine whether the process is in control. Explain your reasoning.

18. **BLOOD TYPES** The distributions of blood types of U.S. and Canadian citizens are shown.

U.S.		Canada	
Type	Distribution	Type	Distribution
O	44%	O	46%
A	42%	A	42%
B	10%	B	9%
AB	4%	AB	3%

a. If 50 U.S. citizens are selected at random, find the probability that fewer than 20 of those chosen will have type O blood.

b. Find the probability that out of 100 randomly selected Canadian citizens, between 80 and 90 of those chosen will have types O or A blood.

c. What is the probability that two randomly chosen people from the U.S. or Canada will have the same blood type?

H.O.T. Problems Use Higher-Order Thinking Skills

19. **ERROR ANALYSIS** Weston and Hannah are calculating results for a survey that they are taking as part of a summer internship. They found that of the residents they surveyed, 65% do not recycle. Weston found the probability that fewer than 30 out of 50 random residents do not recycle is 18.7%, while Hannah found that it would be 27.7%. Is either of them correct? Explain your reasoning.

20. **WRITING IN MATH** Explain how the Central Limit Theorem can be used to describe the shape, center, and spread of a distribution of sample means.

21. **CHALLENGE** In the United States, 7% of the male population and 0.4% of the female population are color-blind. Suppose random samples of 100 men and 1500 women are selected. Is there a greater probability that the men's or women's sample will include at least 10 people who are color-blind? Explain your reasoning.

22. **OPEN ENDED** Give an example of a population and a sample of the population. Explain what is meant by the corresponding sampling distribution.

REASONING **Determine whether each statement is *true* or *false*. Explain your reasoning.**

23. As the number of samples increases, a sampling distribution of sample means will approach the normal distribution.

24. In a binomial distribution, $P(X \geq c) \neq P(X > c)$.

25. **WRITING IN MATH** Explain why the normal distribution can be used to approximate a binomial distribution, what conditions are necessary to do so, and why a correction for continuity is needed.

Spiral Review

26. COMMUNITY SERVICE A recent study of 1286 high school seniors revealed that the students completed an average of 38 hours of volunteer work over the summer with a standard deviation of 6.7 hours. Determine the number of seniors who completed more than 42 hours of community service. Assume that the variable is normally distributed. (Lesson 11-3)

27. GAMES Managers of a fitness club randomly surveyed 56 members and recorded the number of days that each member attended the club in a given week. Use the frequency distribution shown to construct a probability distribution for the random variable X. Then find the mean, variation, and standard deviation of the probability distribution. (Lesson 11-2)

Days, X	Frequency	Days, X	Frequency
0	3	4	11
1	5	5	9
2	10	6	3
3	14	7	1

Find each sum. (Lesson 10-2)

28. $\sum_{n=1}^{19} -50 + 5n$

29. $\sum_{n=12}^{68} 5 - \frac{n}{4}$

30. $\sum_{n=10}^{16} 24n - 90$

Find the specified term of each sequence. (Lesson 10-1)

31. 7th term, $a_n = (a_{n-1} - 6)^2, a_1 = 4$

32. 6th term, $a_n = 3n^2 - 4n$

33. 4th term, $a_n = (a_{n-1})^2 - 11, a_1 = 3$

Find rectangular coordinates for each point with the given polar coordinates. (Lesson 9-3)

34. $\left(2, \frac{\pi}{2}\right)$

35. $\left(\frac{1}{4}, \frac{\pi}{4}\right)$

36. $(6, 210°)$

Find each of the following for $\mathbf{p} = \langle 4, 0 \rangle$, $\mathbf{q} = \langle -2, -3 \rangle$, and $\mathbf{t} = \langle -4, 2 \rangle$. (Lesson 8-2)

37. $\mathbf{p} - \mathbf{t} - 2\mathbf{q}$

38. $\mathbf{q} - 4\mathbf{p} + 3\mathbf{t}$

39. $4\mathbf{p} + 3\mathbf{q} - 6\mathbf{t}$

Write an equation for and graph each parabola with focus F and the given characteristics. (Lesson 7-1)

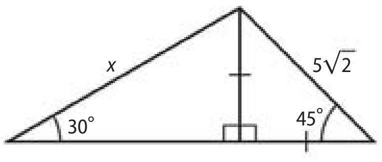
40. $F(-6, 8)$; opens up; contains $(0, 16)$

41. $F(2, -5)$; opens down; contains $(10, -11)$

42. YOGURT The Frozen Yogurt Shack sells cones in three sizes: small, \$2.89; medium, \$3.19; and large, \$3.39. On Friday, 78 cones were sold totaling \$246.42. The Shack sold six more medium cones than small cones that day. Use Cramer's Rule to determine the number of each type of cone sold on Friday. (Lesson 6-3)

Skills Review for Standardized Tests

43. SAT/ACT What is the value of x?



A $2\sqrt{2}$
B 5
C $5\sqrt{3}$
D 10
E $5\sqrt{6}$

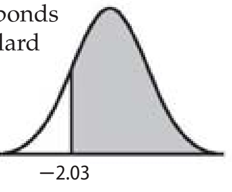
44. REVIEW In a study, 62% of registered voters said they voted in the 2008 presidential election. If 6 registered voters are chosen at random, what is the probability that at least 4 of them voted?

F 32%
G 41.2%
H 58.6%
J 73.2%

45. The average number of patients who are seen every week at a certain hospital is normally distributed. The average per week is 12,423, with a standard deviation of 3269. If a week is selected at random, find the probability that there would be fewer than 4000 patients.

A 0.50%
B 2.37%
C 32.20%
D 36.73%

46. REVIEW Find the area that corresponds to the shaded region of this standard normal distribution.



F 0.02
G 0.04
H 0.96
J 0.98

CHAPTER 11

Mid-Chapter Quiz

Lessons 11-1 through 11-4

1. AUDITIONS The ages of 20 students who auditioned for roles in a high school production of *Gone with the Wind* are shown. (Lesson 11-1)

Ages of Students				
14	15	17	16	14
16	17	16	18	16
15	16	18	15	17
14	18	15	17	16

a. Construct a histogram, and use it to describe the shape of the distribution.

b. Summarize the center and spread of the data using either the mean and standard deviation or the five-number summary. Justify your choice.

2. VACATION Angie is planning a trip for spring vacation. She has narrowed her choices down to two locations. The temperatures during twelve days around the time of spring vacation are shown below for each location. (Lesson 11-1)

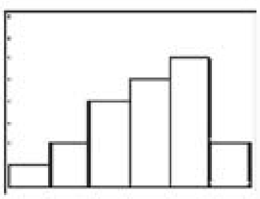
Cape Hatteras, North Carolina					
52	60	62	57	55	63
64	59	54	52	54	60
Orlando, Florida					
77	77	76	76	72	71
72	74	74	72	73	73

a. Construct side-by-side box plots of the data sets, and use this display to compare the center and spread of the distributions.

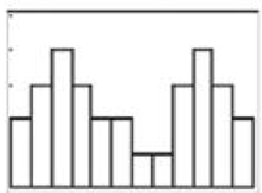
b. Which location has the greater variation in temperature?

3. MULTIPLE CHOICE Which of the following displays a data set that is positively skewed? (Lesson 11-1)

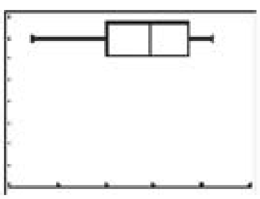
A



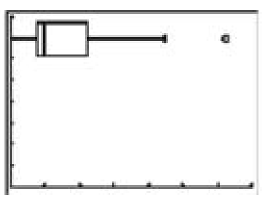
C



B



D



Classify each random variable X as *discrete* or *continuous*. Explain your reasoning. (Lesson 11-2)

4. X represents the number of times a coin lands on heads if flipped a random number of times.

5. X represents the time it takes for a randomly chosen marathon runner to complete a race.

6. TRAVEL In a recent poll, 20% of teenagers said that they have visited Washington, D.C. Find the probability that out of 6 randomly chosen teenagers, at least 3 have visited the nation's capital. (Lesson 11-2)

7. SHAMPOO The amount of water in milliliters in a particular shampoo is normally distributed with $\mu = 125$ and $\sigma = 7$. Find each of the following. (Lesson 11-3)

a. $P(X < 105)$

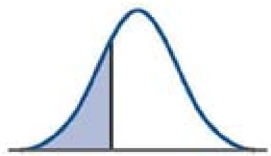
b. $P(X > 140)$

c. $P(115 < X < 130)$

8. GOLF A random sample of 130 golfers resulted in an average score of 78 with a standard deviation of 6.3. Find the number of golfers with an average of 70 or lower. (Lesson 11-3)

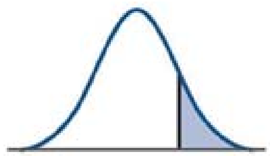
Find the area that corresponds with the shaded region. (Lesson 11-3)

9.



-0.62

10.



1.5

11. PROJECTS The scores on a science project for one class are normally distributed with $\mu = 78$ and $\sigma = 8$. Find each probability. (Lesson 11-3)

a. $P(X \geq 96)$

b. $P(60 < X < 85)$

Find the probability of each sample mean. (Lesson 11-4)

12. $P(\bar{x} < 38)$; $\mu = 40$, $\sigma = 5.5$, $n = 25$

13. $P(\bar{x} > 82.2)$; $\mu = 82.5$, $\sigma = 4.1$, $n = 50$

14. EMPLOYMENT According to a recent study, the average age that a person starts his or her first job is 16.8 years old. Assume that this variable is normally distributed with a standard deviation of 1.7 years. If a random sample of 25 people is selected, find the probability that the mean age the participants started their first jobs is greater than 17 years old. (Lesson 11-4)

LESSON 11-5 Confidence Intervals

::Then

- You analyzed sample means and the effect of the Central Limit Theorem on a sampling distribution. (Lesson 11-4)

::Now

- **1** Use the normal distributions to find confidence intervals for the mean.
- **2** Use *t*-distributions to find confidence intervals for the mean.

::Why?

- Executives at a film studio want to know the average age of people seeing a movie. A survey of 200 people who saw the movie finds that the average age was 20.4 years. The studio executives decide to estimate the mean age *a* for all customers as between 18.1 and 22.7 or $18.1 < a < 22.7$.

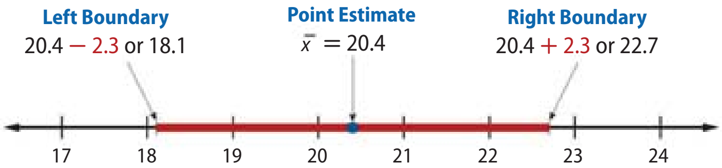
NewVocabulary

inferential statistics
parameter
point estimate
interval estimate
confidence level
maximum error of estimate
critical value
confidence interval
t-distribution
degrees of freedom

1 Normal Distribution

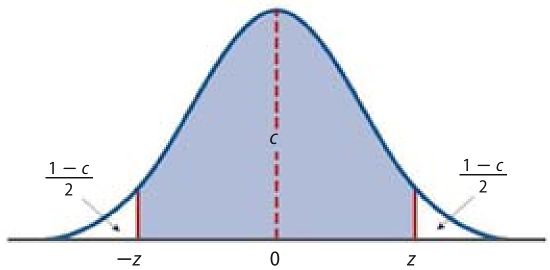
In **inferential statistics**, a sample of data is analyzed and conclusions are made about the entire population. This procedure is used because it is usually too challenging to get information from every member of a population. A measure that describes a characteristic of a population, such as the mean or standard deviation, is called a **parameter**. Many different parameters may be used to analyze data; but in this lesson, we will concentrate on the mean.

The average age of 20.4 years is an example of a **point estimate**, a single value estimate of an unknown population parameter. Due to sampling error and relatively small sampling size, the point estimate will most likely not match the population mean. For this reason, the studio executives used an **interval estimate** of $18.1 < a < 22.7$. An interval estimate is a range of values used to estimate an unknown population parameter. To form an interval estimate, a point estimate is used as the center of the interval and a margin of error is added to and subtracted from the point estimate. For this study, the studio executives let the margin of error be 2.3 years.



Before an interval estimate is created, it is helpful to know just how reliable you want it to be. The probability that the interval estimate will include the actual population parameter is known as the **confidence level**, and is denoted as *c*. We can illustrate a confidence level using the normal distribution if the standard deviation of the population σ is known and the population is normally distributed or if $n \geq 30$. Recall that the Central Limit Theorem states that when $n \geq 30$, the sampling distribution of sample means resembles a normal distribution.

The confidence level for a normal distribution is equal to the area under the standard normal curve between $-z$ and z, as shown. The remaining area in the two tails is then $1 - c$, or $\frac{1}{2}(1 - c)$ for each tail.



Suppose you are conducting an experiment in which you want a confidence level of 95%. When $c = 95\%$, 2.5% of the area lies to the left of $-z$ and 2.5% lies to the right of z. Using a graphing calculator, you can find the corresponding value for $-z$ to be -1.96 and z to be 1.96. By calculating the product of the z-values and the standard deviation of the sample means $\sigma_{\bar{x}}$, the **maximum error of estimate** E, the maximum difference between the point estimate and the actual value of the parameter, can be determined.

StudyTip

Maximum Error of Estimate The maximum error of estimate E will be a positive value since it is the maximum difference between the point estimate and the actual value of the parameter.

KeyConcept Maximum Error of Estimate

The maximum error of estimate E for a population mean μ is given by

$$E = z \cdot \sigma_{\bar{x}} \text{ or } z \cdot \frac{\sigma}{\sqrt{n}},$$

where z is a critical value that corresponds to a particular confidence level, and $\sigma_{\bar{x}}$ or $\frac{\sigma}{\sqrt{n}}$ is the standard deviation of the sample means. When $n \geq 30$, s, the sample standard deviation, may be substituted for σ.

The z-values that correspond to a particular confidence level are known as **critical values**. The three most widely used confidence levels and their corresponding critical values are shown below.

Confidence Level	z-Value
90%	1.645
95%	1.960
99%	2.576

Example 1 Find Maximum Error of Estimate

TEXTBOOKS A poll of 75 randomly selected college students showed that the students spent an average of \$230 on textbooks per session. Assume from past studies that the standard deviation is \$55. Use a 99% confidence level to find the maximum error of estimate for the amount of money spent by students on textbooks.

In a 99% confidence interval, 0.5% of the area lies in each tail. You can find the corresponding z-value to be 2.576 by using a graphing calculator or the table above.

$E = z \cdot \frac{\sigma}{\sqrt{n}}$ — Maximum Error of Estimate

$= 2.576 \cdot \frac{55}{\sqrt{75}}$ — $z = 2.576$, $\sigma = 55$, and $n = 75$

≈ 16.36 — Simplify.

This means that you can be 99% confident that the population mean of money spent on textbooks will be no more than \$16.36 from the sample mean of \$230.

GuidedPractice

1. **MUSIC** Executives at a music label surveyed 125 people who actively download music and found that the listeners have an average of 740 MP3s downloaded to their computers. Assume a standard deviation of 86 MP3s. Use a 94% confidence level to find the maximum error of estimate for the amount of MP3s on the computer of someone who actively downloads music.

Once a confidence level is established and a maximum error of estimate is calculated, it can be used to determine a **confidence interval**. A confidence interval, denoted CI, is a specific interval estimate of a parameter and can be found when the maximum error of estimate is added to and subtracted from the sample mean.

KeyConcept Confidence Interval for the Mean

A confidence interval CI for a population mean μ is given by

$$CI = \bar{x} \pm E \text{ or } \bar{x} \pm z \cdot \frac{\sigma}{\sqrt{n}},$$

where $\bar{x}$ is the sample mean and E is the maximum error of estimate.

Example 2 Find Confidence Intervals When σ is Known

HOMEWORK **A poll of 20 randomly selected high school students showed that the students spent a mean time of 35 minutes per weeknight on homework. Assume a normal distribution with a standard deviation of 12 minutes. Find a 90% confidence interval for the mean of all of the students.**

Substitute 1.645, the corresponding z-value for a 90% confidence interval, into the confidence interval formula.

$CI = \bar{x} \pm z \cdot \frac{\sigma}{\sqrt{n}}$	Confidence Interval for the Mean
$= 35 \pm 1.645 \cdot \frac{12}{\sqrt{20}}$	$\bar{x} = 35$, $z = 1.645$, $\sigma = 12$, and $n = 20$
$\approx 35 \pm 4.41$	Simplify.

Add and subtract the margin of error.

Left Boundary	Right Boundary
$35 - \mathbf{4.41} = 30.59$	$35 + \mathbf{4.41} = 39.41$

A 90% confidence interval is $30.59 < \mu < 39.41$. Therefore, with 90% confidence, the mean time spent on homework by students is between 30.6 and 39.4 minutes.

Guided Practice

2. **SHOPPING** A sample of 65 randomly selected mall patrons showed that they spent an average of \$70 that day. Assume a standard deviation of \$12. Find a 95% confidence interval for the average amount spent by a mall patron that day.

Real-World Career

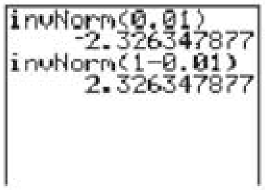
Engineering Engineers use science and math to find economical solutions to technical problems. A bachelor's degree in engineering is usually required for most entry-level jobs.

In many real-world situations, the population standard deviation is unknown. When this is the case, the standard deviation s of the sample can be used in place of the population standard deviation, as long as the variable is normally distributed and $n \geq 30$.

Real-World Example 3 Find Confidence Intervals When σ is Unknown

ENGINEERING **Tensile strength is the stress at which a material breaks or deforms. A company wants to estimate the mean tensile strength of a new material. A random sample of 40 units is normally distributed with an average tensile strength of 36.3 ksi, or 36,300 pounds per square inch, and a standard deviation of 2.9 ksi. Find a 98% confidence interval for the mean tensile strength of the material.**

In a 98% confidence interval, 1% of the area lies in each tail. You can find the corresponding z-value to be 2.33 by using a graphing calculator.



Since the distribution is normal and $n \geq 30$, the sample standard deviation can be used to find a confidence interval.

$CI = \bar{x} \pm z \cdot \frac{s}{\sqrt{n}}$	Confidence Interval for the Mean
$= 36.3 \pm 2.33 \cdot \frac{2.9}{\sqrt{40}}$	$\bar{x} = 36.3$, $z = 2.33$, $s = 2.9$, and $n = 40$
$\approx 36.3 \pm 1.06$	Simplify.

Therefore, a 98% confidence interval is $35.2 < \mu < 37.4$.

Guided Practice

3. **ENGINEERING** Suppose a random sample of 50 units of the same material is normally distributed with an average tensile strength of 39.2 ksi and a standard deviation of 3.1 ksi. Estimate the mean tensile strength with 99% confidence.

Technology Tip

Calculate Confidence Intervals You can check your answer by using a graphing calculator. Press STAT and select ZInterval under the TESTS menu. For Input: select Stats and then enter each of the values. Then select Calculate.

Nick Hanna/Alamy

2 *t*-Distribution

In many cases, the population standard deviation is not known and, due to constraints such as time and cost, sample sizes exceeding 30 are not realistic. In these cases, another distribution called the t-distribution can be used, as long as the variable is approximately normally distributed.

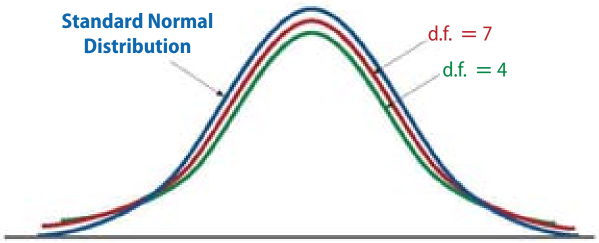
StudyTip

Distributions That Are Not Normal You cannot use a normal distribution or t-distribution to construct a confidence interval if the population is not normally or approximately normally distributed.

The **t-distribution** is a family of curves that are dependent on a parameter known as the *degrees of freedom*. The **degrees of freedom (d.f.)** are equal to $n - 1$ and represent the number of values that are free to vary after a sample statistic is determined.

For example, if $\bar{x} = 4$ in a sample of 10 values, 9 of the 10 values are free to vary. Once the first 9 values are selected, the tenth value must be a specific number in order for $\bar{x} = \frac{40}{10}$. So, the degrees of freedom are $10 - 1$ or 9, which corresponds to a specific curve.

Notice in the figure below that as the degrees of freedom increase, or as d.f. approaches 30, the t-distribution approaches the standard normal distribution.



The characteristics of the t-distribution are summarized below.

KeyConcept Characteristics of the *t*-Distribution

- The distribution is bell-shaped and symmetric about the mean.
- The mean, median, and mode equal 0 and are at the center of the distribution.
- The curve never touches the x-axis.
- The standard deviation is greater than 1.
- The distribution is a family of curves based on the sample size n.
- As n increases, the distribution approaches the standard normal distribution.

Similar to the normal distribution, the t-distribution can be used to construct a confidence interval by using a t-value rather than a z-value to calculate the maximum error of the estimate E. A t-value can be found by

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} \text{ or } \frac{\bar{x} - \mu}{\sigma_{\bar{x}}}, \text{ where } \mu \text{ is the population mean.}$$

You will use a graphing calculator to find values for t since the population mean μ is the parameter that you are attempting to estimate. You can find a confidence interval when using the t-distribution by using the formula shown.

StudyTip

Distributions When $n \geq 30$, it is standard procedure to use the normal distribution. However, t-distributions can still be used.

KeyConcept Confidence Interval Using *t*-Distribution

A confidence interval CI for the t-distribution is given by

$$CI = \bar{x} \pm t \cdot \frac{s}{\sqrt{n}},$$

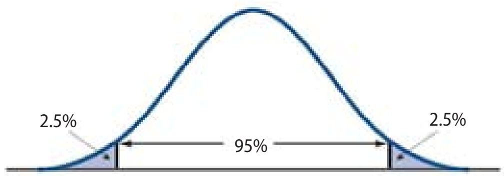
where $\bar{x}$ is the sample mean, t is a critical value with $n - 1$ degrees of freedom, s is the sample standard deviation, and n is the sample size.

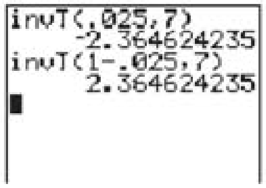
Example 4 Find Confidence Intervals with the t-Distribution

CAPACITY The capacities of eight randomly selected tanks are measured. The mean capacity is 143 liters and the standard deviation is 3.0. Find the 95% confidence interval of the mean capacity of the tanks. Assume that the variable is normally distributed.

The population standard deviation is unknown and $n < 30$, so the t-distribution must be used. Since $n = 8$, there are $8 - 1$ or 7 degrees of freedom. You can use a graphing calculator to find the corresponding t-value.

In the DISTR Menu, select InvT(a, df). The a value represents the area of one tail of the distribution and df represents degrees of freedom. So, for a 95% confidence interval, the area in either tail of the t-distribution is half of 5% or 0.025.





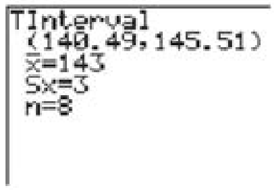
$CI = \bar{x} \pm t \cdot \frac{s}{\sqrt{n}}$ — Confidence Interval Using t-Distribution

$= 143 \pm 2.365 \cdot \frac{3}{\sqrt{8}}$ — $\bar{x} = 143$, $t = 2.365$, $s = 3$, and $n = 8$

$= 143 \pm 2.5$ — Simplify.

Therefore, the 95% confidence interval is $140.5 < \mu < 145.5$.

CHECK You can check your answer by using a graphing calculator. In the STAT menu, select TESTS and TInterval. Under the TInterval menu, select Stats and enter each of the values. Then select Calculate.



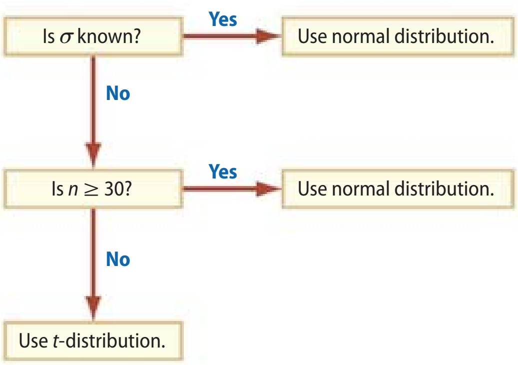
Guided Practice

4. RESTAURANTS The waiting time of ten randomly selected customers at a restaurant was measured with a mean of 25 minutes and a standard deviation of 4 minutes. Find the 99% confidence interval of the mean waiting time for all customers, assuming that the variable is normally distributed.

StudyTip

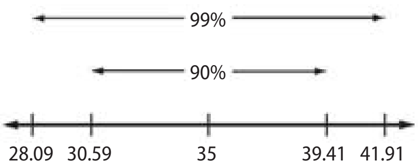
Using the t-Distribution Most real-world inferences about the population mean will be completed using the t-value because σ is rarely known.

It may be difficult to determine whether to use a normal distribution or a t-distribution for a given problem. The chart shown below summarizes when to use each, assuming that the population is normally or approximately normally distributed.



In all statistical experiments, the user determines the confidence level, which directly affects the confidence interval. With all other variables held constant, increasing the confidence level will expand the confidence interval. Expanding a confidence interval reduces the accuracy of the estimate. For example, observe the confidence interval from Example 2 when the confidence level is raised to 99%.

	90% Confidence Level	99% Confidence Level
z-value	1.645	2.576
E	4.41	6.91
CI	$30.59 < \mu < 39.41$	$28.09 < \mu < 41.91$



Generally, a high confidence level and a small confidence interval are desired. This can be achieved by increasing the sample size n. You can find the minimum sample size needed for a specific maximum error of estimate by starting with the formula for E and solving for n.

$$E = z \cdot \frac{\sigma}{\sqrt{n}} \quad \text{Maximum Error of Estimate}$$

$$\sqrt{n} \cdot E = z \cdot \sigma \quad \text{Multiply each side by } \sqrt{n}.$$

$$\sqrt{n} = \frac{z \cdot \sigma}{E} \quad \text{Divide each side by } E.$$

$$n = \left(\frac{z\sigma}{E}\right)^2 \quad \text{Square each side.}$$

KeyConcept Minimum Sample Size Formula

The minimum sample size needed when finding a confidence interval for the mean is given by $n = \left(\frac{z\sigma}{E}\right)^2$, where n is the sample size and E is the maximum error of estimate.

Example 5 Find a Minimum Sample Size

PRODUCT DEVELOPMENT **You are testing the reliability of a thermometer. Your manager asks you to conduct an experiment with results that are accurate to ±0.05 degree with 95% confidence. If $\sigma = 0.8$, how many measurements are needed?**

$$n = \left(\frac{z\sigma}{E}\right)^2 \quad \text{Minimum Sample Size Formula}$$

$$= \left(\frac{1.96(0.8)}{0.05}\right)^2 \quad z = 1.96,\ \sigma = 0.8, \text{ and } E = 0.05$$

$$= 983.45 \quad \text{Simplify.}$$

At least 984 observations are needed to have a margin of error of ±0.05 with 95% confidence.

StudyTip

Rounding It is not possible to have a sample size that is a fraction. Therefore, when finding a minimum sample size, always round answers in the form of a fraction or a decimal to the next greater whole number.

Guided Practice

5. **MARKETING** Executives at a car dealership want to estimate the average age of their customers before making a television commercial. They want to be 90% confident that the mean age is ±2 years of the sample mean. If the standard deviation from a previous study is 12 years, how large should the sample be?

Exercises

= Step-by-Step Solutions begin on page R29.

1. **TRANSPORTATION** A random sample of 85 New York City residents showed that the average commuting time to work was 36.5 minutes. Assume that the standard deviation from previous studies was 11.3 minutes. Find the maximum error of estimate for a 99% confidence level. Then create a confidence interval for the mean commuting time of all New York City residents. (Examples 1–3)

2. **ORANGES** The owner of an orange grove randomly selects 50 oranges of the same type and weighs them with a resulting mean weight of 7.45 ounces and a standard deviation of 0.8 ounce. Find the maximum error of estimate for a 98% confidence level. Then estimate the mean weight of the oranges using a confidence interval. (Examples 1–3)

3. **TEMPERATURE** The average body temperature for 15 randomly selected polar bears was 97.5°F. Assume that the standard deviation from a recent study was 2.8°F. Find the maximum error of estimate for a 95% confidence level. Then estimate the mean body temperature for all polar bears in that region using a confidence interval. (Examples 1–3)

4. **TYPING SPEED** In a random sample of 20 students in a computer class, the average keyboarding speed was 40 words per minute (WPM) with a standard deviation of 8 WPM. Estimate the mean keyboarding speed for all students taking the class using a 90% confidence level. (Example 4)

5. **TEXT MESSAGES** A random sample of 25 students with cell phones found on average, the students send or receive 68 text messages a day with a standard deviation of 13 messages. Estimate the mean number of text messages for all students with cell phones using a 96% confidence level. (Example 4)

6. **COLLEGE VISITS** A random sample of 20 college-bound juniors found on average they visited 6.4 colleges with a standard deviation of 1.9. Estimate the mean number of college visits for all college-bound juniors using a 95% confidence level. (Example 4)

Determine whether the normal distribution or *t*-distribution should be used for each question. Then find each confidence interval given the following information. (Examples 2–4)

7. 90%; $\bar{x} = 128$, $s = 7$, $n = 20$
8. 95%; $\bar{x} = 65$, $s = 15.9$, $n = 300$
9. 95%; $\bar{x} = 39.4$, $s = 1.2$, $n = 15$
10. 98%; $\bar{x} = 122.3$, $\sigma = 2.2$, $n = 2000$
11. 99%; $\bar{x} = 28.3$, $\sigma = 4.5$, $n = 75$
12. 99%; $\bar{x} = 2489$, $\sigma = 18.3$, $n = 160$

13. **COFFEE** The owner of a coffee shop wants to determine the average price for a small cup of coffee in his city. How large should the sample be if he wishes to be accurate to within \$0.015 at 90% confidence? A previous study showed that the standard deviation of the price was \$0.10. (Example 5)

14. **TESTS** A teacher wants to estimate the average amount of time it takes students to finish a 25-question test. How large should the sample be if the teacher wishes to be 99% accurate within 8 minutes? A previous study showed that the standard deviation of the time was 11.3 minutes. (Example 5)

15. **SCHOOL** A survey was taken by 26 randomly selected students, recording the amount of time each student participated in after-school activities for a given week. Assume that the time is normally distributed.

Time (hours)						
11	7	2	7	6	12	9
10	8	6	4	8	8	7
4	7	8	8	6	5	
9	9	10	15	12	13	

   a. Decide the type of distribution that can be used to estimate the population mean. Explain your reasoning.
   b. Calculate the mean and the standard deviation to the nearest tenth.
   c. Construct a 95% confidence interval for the average amount of time students participate in after-school activities.
   d. Interpret the confidence interval in the context of the problem.

16. **WAGES** In a previous study, the standard deviation for starting wages among employed high school students was \$0.50. A survey of 20 randomly selected employed high school students was conducted and their starting wages were recorded. Assume that the wages are normally distributed.

Wages (\$)				
6.75	6.50	6.50	5.50	6.75
5.75	6.50	7.50	7.25	6.00
6.50	7.25	6.75	6.00	5.75
6.00	6.50	6.75	7.00	6.25

   a. Decide the type of distribution that can be used to estimate the population mean. Explain your reasoning.
   b. Calculate the mean to the nearest hundredth.
   c. Construct a 90% confidence interval for the average starting wage for an employed high school student.
   d. Interpret the confidence interval in the context of the problem.

17. **AGE** Julian wants to estimate the average age of teachers with a 95% confidence level. He knows that the standard deviation from past studies is 9 years. If Julian has only 50 teachers at his school to survey, how accurate can he make his estimate?

18. **TELEVISION** Felix and Tanya want to compare the average amount of time per day in minutes that boys and girls watch television. They surveyed 16 female students and 16 male students chosen at random and recorded the viewing times.

Female	
115	120
125	130
120	120
125	105
110	115
105	110
120	125
110	115

Male	
90	140
120	110
105	115
125	120
105	130
150	125
120	110
115	90

a. Calculate the mean and sample standard deviation for each data set.

b. Construct two 99% confidence intervals for the average amount of time spent watching television for both boys and girls.

c. Make a statement comparing the effectiveness of the two intervals.

19. **RESTAURANT** The owner of a restaurant wants a mean preparation time of 20 minutes for each order that is placed. To help ensure that the goal is achieved, the owner timed 24 randomly selected orders and found an average preparation time of 22 minutes with a standard deviation of 4 minutes. The owner will be satisfied if the goal preparation time falls within the 99% confidence interval at which the restaurant is currently operating. Is the owner satisfied? Explain your reasoning.

20. **INCOME** Diego is being transferred by his employer and has his choice of three cities. Before making a decision, he wants to compare the average incomes of his fellow employees in the three cities. With help from the human resource department, he records the following information. The sample size for each city was 35 employees.

City	$\bar{x}$ (\$)	σ (\$)
1	46,700	6300
2	47,800	3000
3	45,000	8000

a. Construct a 95% confidence interval for the average income of employees in each city.

b. If salary is all that is considered, to what city should Diego choose to be transferred? Explain your reasoning.

21. **CELL PHONES** A cell phone manufacturer wants the mean talk time, the time a phone is engaged in sending a message or transmitting a conversation, for its long-life batteries to be 62 hours. To ensure the quality of its batteries, the manufacturer randomly samples 14 phones and records the talk times in hours.

Talk Time (hours)						
61.0	63.1	63.3	59.1	63.4	61.5	60.0
62.6	62.3	60.3	62.9	61.3	62.4	63.6

The manufacturer will be satisfied if the mean talk time falls within the 99% confidence interval at which the batteries are currently operating. Is the manufacturer satisfied? Explain your reasoning.

22. **READING** Rama is conducting a study on the average amount of time that people between the ages of 17–25 spend reading each day. She surveys 20 people chosen at random. She currently has a confidence interval with a maximum error of estimate of 10 minutes. What will Rama's sample size need to be if she wants to reduce the error to 5 minutes? 2.5 minutes?

H.O.T. Problems Use Higher-Order Thinking Skills

23. **CHALLENGE** A confidence interval of $40.872 < \mu < 49.128$ is created by a random study. If the sample standard deviation was 10 and the t-value used was 2.064, find the degrees of freedom for the t-distribution.

24. **WRITING IN MATH** Most studies desire results with high confidence levels. Explain why a 99% confidence level is not used for every study.

REASONING **Determine whether each statement is *true* or *false*. Explain your reasoning.**

25. Increasing the sample size will expand a confidence interval.

26. Increasing the confidence level will expand a confidence interval.

27. Increasing the standard deviation will expand a confidence interval.

28. Increasing the mean will expand a confidence interval.

29. **REASONING** If a person conducting a study wants to reduce the maximum error of estimate by $\frac{1}{x}$, what would the person have to do to the sample size?

30. **CHALLENGE** A study conducted with a random sample of size $n = 64$ results in a confidence interval of $3.19 < \mu < 4.01$. If the interval was created using a 90% confidence level, find the sample standard deviation.

31. **WRITING IN MATH** Explain why parameters are needed in statistics.

Spiral Review

32. EDUCATION According to a recent survey, 35% of adult Americans had obtained a bachelor's degree. What is the probability that in a random sample of 50 people, between 12 and 16 people had earned a bachelor's degree? (Lesson 11-4)

33. CAR BATTERY The useful life of a certain car battery is normally distributed with a mean of 100,000 miles and a standard deviation of 10,000 miles. The company makes 20,000 batteries a month. What is the probability that if you select a car battery at random, it will last between 80,000 and 110,000 miles? (Lesson 11-3)

Use the fifth partial sum of the trigonometric series for cosine or sine to approximate each value to three decimal places. (Lesson 10-6)

34. $\sin \frac{\pi}{7}$

35. $\cos \frac{2\pi}{11}$

36. $\sin \frac{4\pi}{17}$

Determine the eccentricity, type of conic, and equation of the directrix given by each polar equation. (Lesson 9-4)

37. $r = \dfrac{8}{\cos \theta + 5}$

38. $r = \dfrac{4}{7 \cos \theta + 4}$

39. $r = \dfrac{2}{\sin \theta + 3}$

Use the dot product to find the magnitude of the given vector. (Lesson 8-3)

40. $\mathbf{u} = \langle -8, 0 \rangle$

41. $\mathbf{v} = \langle 7, 2 \rangle$

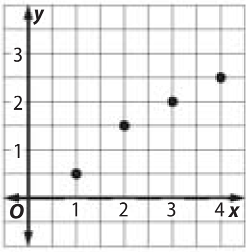
42. $\mathbf{u} = \langle 4, 8 \rangle$

43. THRILL RIDES At an amusement park, there is an additional cost per person to ride the Cloud Coaster as well as the Danger Coaster as a 1-person, 2-person, and 3-person ride. The table shows how many people paid for the rides during the first four hours that the park was open. Write and solve a system of equations to determine the cost of each ride per person. Interpret your solution. (Lesson 6-1)

Hour	Cloud Coaster	Danger Coaster			Total Paid ($)
		1 Person	2 Person	3 Person	
1	8	5	10	3	575
2	10	8	2	6	574
3	16	4	8	3	661
4	13	11	6	0	722

Skills Review for Standardized Tests

44. SAT/ACT Which line best fits the data in the graph?



A $y = x$

B $y = -0.5x + 4$

C $y = -0.5x - 4$

D $y = 0.5 + 0.5x$

E $y = 0.75x$

45. REVIEW People were chosen at random and asked how many times they went out to eat per week. If $\sigma = 0.6$, the results had 95% confidence, and they were accurate to ± 0.05, about how many people were asked?

F 6 **G** 23 **H** 144 **J** 554

46. In a random sample of 28 college-educated adults aged 25 to 35, the average amount of student-loan debt was \$5566 with a standard deviation of \$1831. Estimate the mean student-loan debt for all college-educated adults ages 25 to 35 using a 90% confidence interval.

A $4188 < \mu < 6944$

B $4319 < \mu < 6813$

C $4507 < \mu < 6625$

D $4997 < \mu < 6135$

47. REVIEW A school has two independent backup generators having probabilities of 0.9 and 0.95, respectively, of successful operation in case of a power outage. What is the probability that at least one backup generator operates during a power outage?

F 0.855

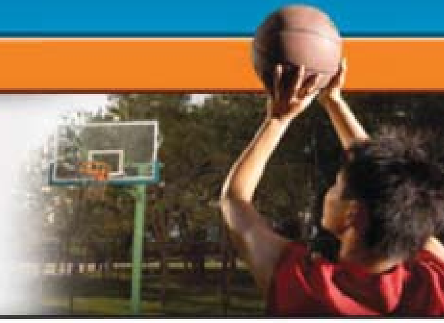
G 0.89

H 0.95

J 0.995

LESSON

11-6 Hypothesis Testing



Then

- You found confidence intervals for the means of distributions. (Lesson 11-5)

Now

1. Write null and alternative hypotheses, and identify which represents the claims.
2. Perform hypothesis testing using test statistics and *p*-values.

Why?

- Luther and Jimmy are shooting baskets when Jimmy proudly boasts, "I can make at least 90% of my free throws." Luther is curious about Jimmy's remark and wants to test the accuracy of his claim.

NewVocabulary
hypothesis test
null hypothesis
alternative hypothesis
level of significance
left-tailed test
two-tailed test
right-tailed test
p-value

1 Hypotheses

A **hypothesis test** assesses evidence provided by data about a claim concerning a population parameter. A claim of this type is called a *statistical hypothesis* and may or may not be true. Jimmy's claim at the beginning of the lesson is an example of a statistical hypothesis.

To test the validity of a claim, write it as a mathematical statement. Jimmy's claim can be written as $\mu \geq 90\%$, where μ is his average shooting percentage. The statement $\mu < 90\%$ is the complement of the original statement, which represents Jimmy not meeting his claim. These two statements represent the pair of hypotheses that need to be stated to test a claim.

- The **null hypothesis** H_0: There *is not* a significant difference between the sample value and the population parameter. It will contain a statement of *equality*, such as $\geq$, $=$, or $\leq$. In this example, $\mu \geq 90\%$ is the null hypothesis.
- The **alternative hypothesis** H_a: There is a difference between the sample value and the population parameter. It will contain a statement of *inequality*, such as $>$, $\neq$, or $<$. In this example, $\mu < 90\%$ is the alternative hypothesis.

If a claim k is made about a population mean μ, the possible combinations for the hypotheses are:

H_0: $\mu = k$ and H_a: $\mu \neq k$ $\quad$ H_0: $\mu \geq k$ and H_a: $\mu < k$ $\quad$ H_0: $\mu \leq k$ and H_a: $\mu > k$

Example 1 Null and Alternative Hypotheses

For each statement, write the null and alternative hypotheses. State which hypothesis represents the claim.

a. Makers of a gum brand claim that their gum will keep its flavor for at least 5 hours.

This claim becomes $\mu \geq 5$ and is the null hypothesis since it includes an equality symbol. The complement is $\mu < 5$.

H_0: $\mu \geq 5$ (Claim) $\quad$ H_a: $\mu < 5$

b. Technicians of an automotive company claim that they will perform an oil change on a car in less than 15 minutes.

This claim becomes $\mu < 15$ and is the alternative hypothesis since it includes an inequality symbol. The complement is $\mu \geq 15$.

H_0: $\mu \geq 15$ $\quad$ H_a: $\mu < 15$ (Claim)

c. A teacher claims that the average amount of time that his students are spending on homework every night is 35 minutes.

This claim becomes $\mu = 35$ and is the null hypothesis since it includes an equality symbol. The complement is $\mu \neq 35$.

H_0: $\mu = 35$ (Claim) $\quad$ H_a: $\mu \neq 35$

GuidedPractice

1A. A football player claims that he can achieve more than 100 rushing yards per game.

1B. A cross country star claims that it will take her no more than 4 minutes to run a mile.

1C. A salesperson claims that she averages 12 sales per month.

WatchOut!

Reject or Not Reject The null hypothesis is the hypothesis tested but it may not represent the claim. For example, if the alternative hypothesis represents the claim and the null hypothesis is rejected, then the claim is actually being supported.

2 Significance and Tests

To validate a claim, the null hypothesis is always tested. In the example at the beginning of the lesson, $\mu \geq 90\%$ would be tested. After a sample of data is analyzed, one of two decisions is made.

- Reject the null hypothesis.
- Do not reject the null hypothesis.

Every shot that Jimmy could ever take cannot be recorded. Luther can only analyze a sample of data, such as having Jimmy take 100 shots. Thus, there is always a chance that Luther can make the wrong decision. When the decision is incorrect, it is either a *Type I* or a *Type II* error.

	H_0 is True	H_0 is False
H_0 is rejected	**Type I Error** The null hypothesis is rejected, when it is actually true. Luther rejects the statement $\mu \geq 90\%$ when Jimmy actually shoots 90% or better.	**Correct decision** Luther rejects the statement $\mu \geq 90\%$ when Jimmy actually shoots less than 90%.
H_0 is not rejected	**Correct decision** Luther does not reject the statement $\mu \geq 90\%$ when Jimmy actually shoots 90% or better.	**Type II Error** The null hypothesis is not rejected, when it is actually false. Luther does not reject the statement $\mu \geq 90\%$ when Jimmy actually shoots less than 90%.

This suggests that there are actually four possible outcomes when a decision about the null hypothesis is made. The only way to guarantee complete accuracy is to test the entire population.

StudyTip

Errors The risk of making a Type I error is identical to the significance level. Reducing the risk of a Type I error *increases* the risk of a Type II error by widening the acceptance interval.

The **level of significance**, denoted α, is the maximum allowable probability of committing a Type I error. For example, if $\alpha = 0.10$, there is a 10% chance that H_0 was rejected when it was actually true, or there is a 90% chance that a correct decision was made. Any level of significance can be chosen. The three most commonly used levels are $\alpha = 0.10$, $\alpha = 0.05$, and $\alpha = 0.01$.

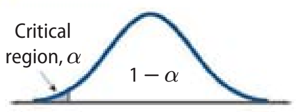
After a level of significance is chosen, a critical value can be found using either a z- or t-value. Similar to confidence intervals, the decision to use a z- or t-value is based on the characteristics of the study.

- If σ is known or $n \geq 30$, use a z-value.
- If σ is unknown and $n < 30$, use a t-value.

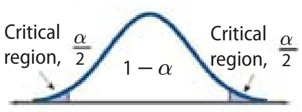
The z- or t-value and the alternative hypothesis will determine the *critical region*, the range of values that suggest a significant enough difference to reject the null hypothesis. The location of a critical region is determined by the inequality sign of the alternate hypothesis, which indicates whether the test is left-tailed, right-tailed, or two-tailed.

ConceptSummary Tests of Significance

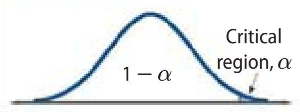
If H_a: $\mu < k$, the hypothesis test is a **left-tailed test**.



If H_a: $\mu \neq k$, the hypothesis test is a **two-tailed test**.



If H_a: $\mu > k$, the hypothesis test is a **right-tailed test**.



Once the area corresponding to the level of significance is determined, a *test statistic* for the sample mean is calculated. The test statistic is the z- or t-value for the sample and will be referred to as the *z statistic* or *t statistic*. If the z or t statistic for the sample:

- is in the critical region, H_0 should be rejected.
- is *not* in the critical region, H_0 should not be rejected.

StudyTip

z Statistic and t Statistic Calculation To calculate a z statistic, use $z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}}$. To calculate a t statistic, use $t = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}}$.

Real-WorldLink

The Nutrition Labeling and Education Act of 1990, or NLEA, required nutrition labeling for most foods, excluding meat and poultry.

Source: U.S. Food and Drug Administration

Robert George Young/Photographer's Choice/Getty Images

The steps to conduct a hypothesis test are summarized below.

KeyConcept Steps for Hypothesis Testing

Step 1 State the hypotheses, and identify the claim.

Step 2 Determine the critical value(s) and region.

Step 3 Calculate the test statistic.

Step 4 Reject or fail to reject the null hypothesis.

Real-World Example 2 One-Sided Hypothesis Test

NUTRITION Representatives of a company report that their product contains no more than 5 grams of fat. A researcher tests a random sample of 50 products and finds that $\bar{x} = 5.03$ grams. If the standard deviation of the population is 0.14 gram, use a 5% level of significance to determine whether there is sufficient evidence to reject the company's claim.

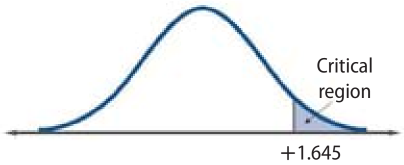
Step 1 State the null and alternative hypotheses, and identify the claim.

The claim written as a mathematical statement is $\mu \leq 5$. This is the null hypothesis. The alternative hypothesis is $\mu > 5$.

H_0: $\mu \leq 5$ (claim) H_a: $\mu > 5$

Step 2 Determine the critical value(s) and region.

The population standard deviation is known and $n \geq 30$, so you can use the z-value. The test is right-tailed since $\mu > 5$. Because a 5% level of significance is called for, $\alpha = 0.05$. Use a graphing calculator to find the z-value.



Step 3 Calculate the test statistic.

Find the z statistic. Since $\sigma = 0.14$ and $n = 50$, $\sigma_{\bar{x}} = \frac{0.14}{\sqrt{50}}$ or about 0.02.

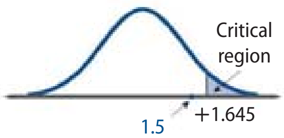
$$z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} \quad \text{Formula for } z \text{ statistic}$$

$$= \frac{5.03 - 5}{0.02} \quad \bar{x} = 5.03,\ \mu = 5, \text{ and } \sigma_{\bar{x}} = 0.02$$

$$= 1.5 \quad \text{Simplify.}$$

Step 4 Reject or fail to reject the null hypothesis.

H_0 is not rejected since the test statistic does not fall within the critical region.



Therefore, there is not enough evidence to reject the claim that there is no more than 5 grams of fat per product.

GuidedPractice

2. **JOBS** Employees at a bookstore claim that the mean wage per hour is less than the competitor's mean wage of \$10.50. If a random sample of 20 employees shows a mean wage of \$10.05 with a standard deviation of \$0.75, test the employees' claim at $\alpha = 0.01$.

For a two-sided test, the level of significance α must be divided by 2 in order to determine the critical value at each tail.

Example 3 Two-Sided Hypothesis Test

FRUIT SNACKS **Representatives of a company have stated that each box of fruit snacks contains 80 pieces. A researcher wants to determine if this is true. A random sample of 25 boxes is selected, with a sample mean of 84.1 pieces and a standard deviation of 7 pieces. Is this statistically significant at $\alpha = 0.01$?**

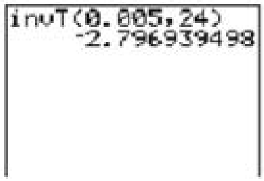
Step 1 **State the null and alternative hypotheses, and identify the claim.**

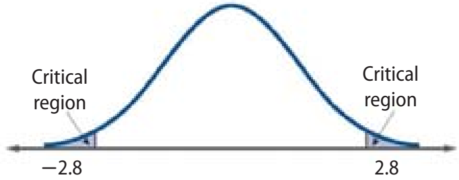
The claim written as a mathematical statement is $\mu = 80$. This is the null hypothesis. The alternative hypothesis is $\mu \neq 80$.

H_0: $\mu = 80$ (claim) H_a: $\mu \neq 80$

Step 2 **Determine the critical value(s) and region.**

The t-value should be used since $n < 30$ and σ is unknown. The test is two-tailed since $\mu \neq 80$, so the critical values are determined by $\frac{\alpha}{2}$ or 0.005. Using a graphing calculator, the critical values for $\alpha = 0.005$ with $25 - 1$ or 24 degrees of freedom are $t = -2.8$ and $t = 2.8$.





WatchOut!

Determining Critical Values

Remember to use the InvT(function to find t-values for the t-distribution and InvNorm(to find z-values for the normal distribution.

Step 3 **Calculate the test statistic.**

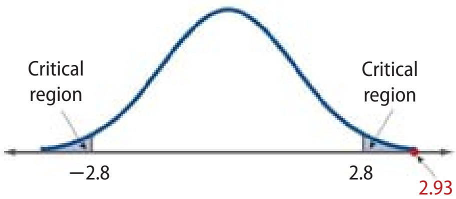
$$t = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} \quad \text{Formula for } t \text{ statistic}$$

$$= \frac{84.1 - 80}{1.4} \quad \bar{x} = 84.1,\ \mu = 80, \text{ and } \sigma_{\bar{x}} = \frac{7}{\sqrt{25}} \text{ or } 1.4$$

$$\approx 2.93 \quad \text{Simplify.}$$

Step 4 **Reject or fail to reject the null hypothesis.**

H_0 is rejected since the test statistic falls within the critical region.



There is enough evidence to reject the claim that there are 80 pieces in each box.

GuidedPractice

3. **TRAVEL** Representatives of a travel bureau in a U.S. city claim that in a recent year, an average of 110 people visited the city every day. In a sample of 90 days, there was an average of 115 visitors per day, with a standard deviation of 18 visitors. At $\alpha = 0.05$, is there enough evidence to reject the claim?

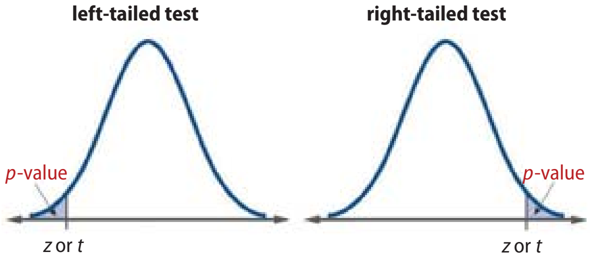
The **p-value** can also be used to determine whether H_0 should be rejected. The p-value is the lowest level of significance at which H_0 can be rejected for a given set of data. After calculating the z or t statistic for a hypothesis test, it can be converted into a p-value using a graphing calculator. To use the p-value to evaluate H_0, compare the p-value to α.

- If $p \leq \alpha$, then reject H_0.
- If $p > \alpha$, then do not reject H_0.

StudyTip

Law of Large Numbers The value of $\bar{x}$ is rarely identical to the true μ and often varies from sample to sample. However, due to the Law of Large Numbers, if we take larger and larger samples, $\bar{x}$ is guaranteed to get closer and closer to the true μ. This is true for *any* distribution.

The *p*-value corresponds to the area found under the normal curve to the left or right of the *z* or *t* statistic calculated for the sample data. The location of the area is determined by the type of test being preformed.



The α value is chosen by the researcher before the statistical test is performed, while the *p*-value is calculated after the sample mean is determined.

Example 4 Hypothesis Testing and *p*-Values

HORTICULTURE A biologist treated 40 plants with a chemical and then compared the amount of growth with untreated plants. For the untreated plants, the mean height is 21.6 centimeters. The treated plants have a mean height of 22.4 centimeters and $s = 1.8$ centimeters. The biologist claims that the chemical increased plant growth. Determine whether this result is significant at $\alpha = 0.01$.

The claim written as a mathematical statement is $\mu > 21.6$. This is the alternative hypothesis. The null hypothesis is $\mu \leq 21.6$.

H_0: $\mu \leq 21.6$ $\qquad$ H_a: $\mu > 21.6$ $\qquad$ (claim)

Since $n \geq 30$, the z statistic is used.

$z = \dfrac{\bar{x} - \mu}{\sigma_{\bar{x}}}$ $\qquad$ Formula for z statistic

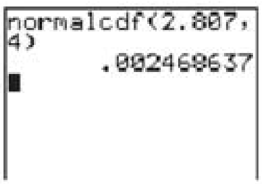
$\approx \dfrac{22.4 - 21.6}{0.285}$ $\qquad$ $\bar{x} = 22.4$, $\mu = 21.6$, and $\sigma_{\bar{x}} = \dfrac{1.8}{\sqrt{40}}$ or about 0.285

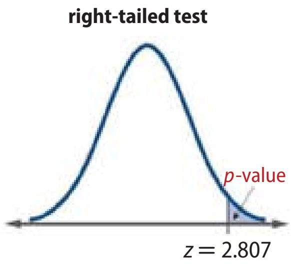
≈ 2.807 $\qquad$ Simplify.

TechnologyTip

Calculating the Area Under a *t*-Curve You can use a graphing calculator to find the area under the *t*-curve that corresponds to any *t*-value by selecting 2nd [DISTR] and tcdf(*lower t value, upper t value, d.f.*).

This is a right-tailed test since the alternative hypothesis is $\mu > 21.6$. The area associated with $z = 2.807$ is 0.0025.





The *p*-value 0.0025 is less than 0.01. Therefore, the null hypothesis is rejected and there is significant evidence that the chemical increased plant growth.

GuidedPractice

4. **DRUGS** Makers of a sleep-aid claim that their product provides more than 8 hours of uninterrupted sleep. In a test of 50 patients, the mean amount of uninterrupted sleep was 8.07 hours with a standard deviation of 0.3 hour. Find the *p*-value and determine if there is enough evidence to reject the claim at $\alpha = 0.03$.

It is important to remember that statistical tests do not prove that a claim is true or false. These types of tests simply state that there is or is not enough evidence to say that a claim is likely to be true.

Exercises

= Step-by-Step Solutions begin on page R29.

Write the null and alternative hypotheses for each statement, and state which hypothesis represents the claim. (Example 1)

1. Makers of a cereal brand claim their product contains 4 grams of fiber.
2. A student claims that he has received at least an 85% on his math tests.
3. Darcy claims that she can drive to school from her house in less than 10 minutes.
4. Vanessa claims that her bowling average is 183.
5. Ian claims that he can recite the names of more than 38 former presidents of the United States.

Calculate the test statistic, and determine whether there is enough evidence to reject the null hypothesis. Then make a statement regarding the original claim.

6. **ADVERTISING** Company representatives claim that they will ship a product in less than four days. If a random selection of 60 delivery times has a sample mean of 3.9 days and a standard deviation of 0.6 day, is there enough evidence to reject the claim at $\alpha = 0.05$? Explain. (Examples 2 and 3)

7. **HEALTH** A researcher claims that a supplement does not increase bone density by at least 0.05 gram per square centimeter. If a study shows that the supplement increased bone density in a random sample of 35 people by 0.048 gram per square centimeter with a standard deviation of 0.004, is there enough evidence to reject the claim at $\alpha = 0.01$? Explain. (Examples 2 and 3)

8. **HOTELS** Owners of a hotel chain claim that the average cost of one of their hotel rooms in the U.S. is \$82. Sample data for 25 hotel rooms is collected. The average cost was found to be \$85 with a standard deviation of \$8. Is there enough evidence to reject the owners' claim at $\alpha = 0.02$? Explain. (Examples 2 and 3)

9. **CALCULATORS** A teacher claims that the average cost of a certain brand of graphing calculator is at least \$90. A random sample of 40 stores shows that the mean cost is \$89.25 with a standard deviation of \$4.95. Is there enough evidence to reject the teacher's claim at $\alpha = 0.05$? Explain. (Examples 2 and 3)

For each claim k, use the specified information to calculate the test statistic and determine whether there is enough evidence to reject the null hypothesis. Then make a statement regarding the original claim. (Examples 2 and 3)

10. k: $\mu = 1240$, $\alpha = 0.05$, $\bar{x} = 1245$, $s = 32$, $n = 50$
11. k: $\mu > 88$, $\alpha = 0.05$, $\bar{x} = 91.2$, $s = 3.9$, $n = 22$
12. k: $\mu < 500$, $\alpha = 0.01$, $\bar{x} = 490$, $s = 27$, $n = 35$
13. k: $\mu \neq 5500$, $\alpha = 0.01$, $\bar{x} = 5430$, $s = 236$, $n = 200$
14. k: $\mu \leq 10{,}000$, $\alpha = 0.01$, $\bar{x} = 10{,}015$, $s = 85$, $n = 18$

15. **REAL ESTATE** A researcher wants to test a claim that the average home sale price in the U.S. is less than \$260,000. She selects a sample of 40 homes and finds the mean sale price of the sample to be \$254,500 with a standard deviation of \$12,500. Find the p-value, and determine whether there is enough evidence to support the claim at $\alpha = 0.05$. (Example 4)

16. **MUSIC** Representatives of an electronics company claim that the average lifetime of an MP3 player is at least 5 years. A random sample of 100 MP3s shows a mean life span of 5.2 years with a standard deviation of 1.2 years. Find the p-value, and determine whether there is enough evidence to support the claim at $\alpha = 0.01$. (Example 4)

17. **BASEBALL** Shelby believes that the cost of attending a baseball game for a family of two adults and two children is under \$125. She surveys 18 families selected at random and finds that the average cost is \$122.88 with a standard deviation of \$13.21. Find the p-value, and determine whether there is enough evidence to support the claim at $\alpha = 0.10$.

18. **CROSS COUNTRY** Pablo claims that the average mile time for the students in his school is less than 7 minutes. He records the mile times of 20 randomly selected students. Determine whether Pablo's claim is supported at $\alpha = 0.05$.

Mile Times (minutes)									
5.25	7.27	5.46	7.63	7.75	5.42	6.00	8.17	9.45	6.20
6.63	7.38	6.97	7.85	7.03	6.53	6.87	7.22	7.16	6.92

a. Write the null and alternative hypotheses, and state which hypothesis represents the claim.
b. Determine whether there is enough evidence to reject the null hypothesis using critical values.
c. Make a statement regarding the original claim.

19. **HOMEWORK** Ms. Taylor claims that her math students spend 25 minutes each night on homework. Ava asks her classmates to record the average amount of time that they spend on homework each night over the course of a week. Determine whether Ms. Taylor's claim is supported at $\alpha = 0.10$.

Times (minutes)											
45	40	10	15	18	20	34	36	20	25	28	25
26	30	22	25	24	29	26	28	23	28	25	26
29	30	22	20	22	24	23	24	25	29	25	

a. Write the null and alternative hypotheses, and state which hypothesis represents the claim.
b. Determine whether there is enough evidence to reject the null hypothesis using critical values.
c. Make a statement regarding the original claim.

Describe the outcome if a type I or a type II error is committed when the null hypothesis is tested.

20. The accused person is not guilty.

21. The *X*-ray came back positive for an ankle sprain.

22. Students use study time efficiently.

23. The majority of students do not have jobs.

24. The average lifespan of a goldfish is 2 years.

25. The venom from the snake is not poisonous.

26. **SLEEP** Mr. King believes that college students get less than 6 hours of sleep each night. He randomly selected a group of students and recorded the average amount of sleep each student gets each night.

Average Sleep (hours/night)								
5.4	6.7	6.5	5.5	5.5	6.0	5.8	6.7	6.8
4.5	5.7	7.5	5.4	5.3	8.0	4.5	4.5	5.0

a. Write the null and alternative hypotheses, and state which hypothesis represents the claim.

b. Find the p-value.

c. Determine whether there is enough evidence to reject the null hypothesis at $\alpha = 0.05$.

d. Make a statement regarding the original claim.

27. **ACT** The average composite score on the ACT is a 21. Instructors of an ACT preparation class claim that they can raise test takers' scores. The scores of randomly selected attendees were recorded.

ACT Scores											
24	23	27	23	19	16	33	30	22	25	23	26
21	30	22	18	28	21	26	32	20	17	23	24
25	28	19	22	21	19	18	20	25	22	24	23

a. Write the null and alternative hypotheses, and state which hypothesis represents the claim.

b. Find the p-value.

c. Determine whether there is enough evidence to reject the null hypothesis at $\alpha = 0.01$.

d. Make a statement regarding the original claim.

28. **QUANTITY** A chocolatier claims that his candy averages at least 81 pieces per bag. Conrad randomly selects bags of the candies and counts the pieces. Assume that $\sigma = 1.2$.

Number of Pieces								
81	80	82	82	83	82	84	81	81
80	83	83	82	81	80	84	81	81

a. Write the null and alternative hypotheses, and state which hypothesis represents the claim.

b. Determine whether there is enough evidence to reject the null hypothesis at $\alpha = 0.05$ using critical values.

c. Find the p-value. Determine whether there is enough evidence to reject the null hypothesis at $\alpha = 0.05$.

29. **GRADES** Mr. Lewis claims that the average grade for his students is an 85%. Two of his students, Victor and Malina, collect the following samples of grades from students in their classes.

Victor's Scores								
64	84	86	99	76	90	79	94	85
84	85	88	91	80	85	76	86	96

Malina's Scores							
95	86	95	83	86	85	84	88
88	86	87	88	95	86	85	95

a. Write the null and alternative hypotheses, and state which hypothesis represents the claim.

b. Suppose $\alpha = 0.10$. For each class, determine whether there is enough evidence to reject the null hypothesis and make a statement regarding the original claim.

c. Determine whether there is enough evidence to reject the null hypothesis if the two samples of data are combined. Use the result to make a statement regarding the original claim.

d. Make a conjecture in regard to the result found in part **c** and the Law of Large Numbers.

H.O.T. Problems Use Higher-Order Thinking Skills

30. **ERROR ANALYSIS** Trish and Molly are completing their statistics homework. Trish claims that it is always better to have a type I error rather than a type II error. Molly disagrees. Is either of them correct? Explain your reasoning.

31. **WRITING IN MATH** Describe the difference between conducting a hypothesis test using test statistics and critical values and conducting a hypothesis test using p-values.

REASONING **Determine whether each statement is *true* or *false*. Explain your reasoning.**

32. If the null hypothesis is rejected, then the claim is always rejected.

33. The alternative hypothesis can include an equality symbol if it represents the claim.

34. The p-value is always going to be a positive value.

35. **CHALLENGE** For a random set of data, $p = 0.0104$, $s = 0.3$, and $\bar{x} = 14.9$. If the study was conducted to test a claim of $\mu < 15$, find n.

36. **WRITING IN MATH** Explain why it may not always be in the researcher's best interest to have the lowest possible significance level in order to reduce the possibility of a type I error.

Spiral Review

37. SHOES A sample of 35 pairs of running shoes found the average cost to be \$45.25 with a standard deviation of \$7.60. Estimate the mean cost for running shoes using a confidence interval given a 90% confidence level. (Lesson 11-5)

38. RARE BOOKS The average prices for three antique books are shown. The prices vary due to the age and condition of each book. (Lesson 11-4)

Book	Average Price (\$)
Don Quixote	155
Moby Dick	98
Oliver Twist	118

a. For a sample of 25 copies of *Don Quixote*, find the probability that the mean price is more than \$160, if the standard deviation is \$18.

b. For a random sample of 40 copies of *Oliver Twist*, find the probability that the mean price will be between \$115 and \$120, if the standard deviation is \$15.

Determine whether each sequence is *convergent* or *divergent*. (Lesson 10-1)

39. $a_n = (10 - n)^2$

40. $a_n = n^2 + 10n - 9$

41. $a_n = \frac{3n + 2}{n}$

Find the angle θ between vectors u and v. (Lesson 8-5)

42. $\mathbf{u} = 2\mathbf{i} + 8\mathbf{j} + 7\mathbf{k}, \mathbf{v} = -7\mathbf{i} - 3\mathbf{j} - 9\mathbf{k}$

43. $\mathbf{u} = -2\mathbf{i} - 8\mathbf{j} + 4\mathbf{k}, \mathbf{v} = 3\mathbf{i} - 4\mathbf{j} - 7\mathbf{k}$

44. $\mathbf{u} = 5\mathbf{i} - 9\mathbf{k}, \mathbf{v} = 2\mathbf{i} - 3\mathbf{j} - 8\mathbf{k}$

Find a unit vector in the same direction as v. (Lesson 8-2)

45. $\mathbf{v} = \langle -9, 2 \rangle$

46. $\mathbf{v} = \langle -5, -1 \rangle$

47. $\mathbf{v} = \langle 4, 3 \rangle$

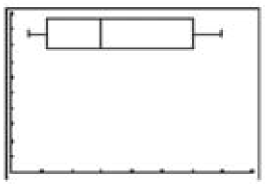
Write an equation for the hyperbola with the given characteristics. (Lesson 7-3)

48. vertices $(-6, 3)$, $(4, 3)$; conjugate axis is 8 units

49. vertices $(-2, 6)$, $(-2, -4)$; conjugate axis is 6 units

Skills Review for Standardized Tests

50. REVIEW Estimate the median and spread of the data represented by the box plot.



[0, 80] scl: 10 by [0, 10] scl: 1

A median ≈ 30, spread ≈ 50

B median ≈ 30, spread ≈ 65

C median ≈ 50, spread ≈ 50

D median ≈ 50; spread ≈ 65

51. REVIEW Find the solutions of $x^2 = i$.

F i and $-i$

G $\cos \frac{\pi}{4} + i \sin \frac{\pi}{4}$ and $\cos \frac{5\pi}{4} + i \sin \frac{5\pi}{4}$

H 1 and −1

J $\cos \frac{\pi}{8} + i \sin \frac{\pi}{8}$ and $\cos \frac{9\pi}{8} + i \sin \frac{9\pi}{8}$

52. SAT/ACT Which of the following statements are true if n is an integer?

I. $3n + 6$ is divisible by 3.

II. $10n + 8$ is divisible by 2.

III. $4n - 2$ is divisible by 4.

A I only

B II only

C I and II only

D I and III only.

E I, II, and III are true.

53. Marcel believes that the average price of gasoline is still under \$2.50 per gallon. He randomly calls 40 different service stations and finds that the average price is \$2.51 with a standard deviation of \$0.06. Find the p-value, and determine whether there is enough evidence to support the claim at $\alpha = 0.10$.

F 0.85; not enough evidence

G 0.95; enough evidence

H 0.15; not enough evidence

J 0.05; enough evidence

LESSON 11-7 Correlation and Linear Regression

Then

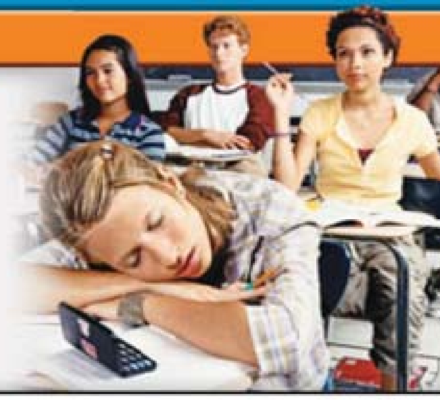
- You analyzed univariate data. (Lessons 11-1 through 11-6)

Now

- 1 Measure the linear correlations for sets of bivariate data using the correlation coefficient, and determine if the correlations are significant.
- 2 Generate least-squares regression lines for sets of bivariate data, and use the lines to make predictions.

Why?

- A feature writer for a school newspaper is interested in determining whether the number of hours of sleep a student gets each night is related to his or her overall grade point average. In statistical terms, the writer would like to know if there is a *correlation* between sleep and grades.



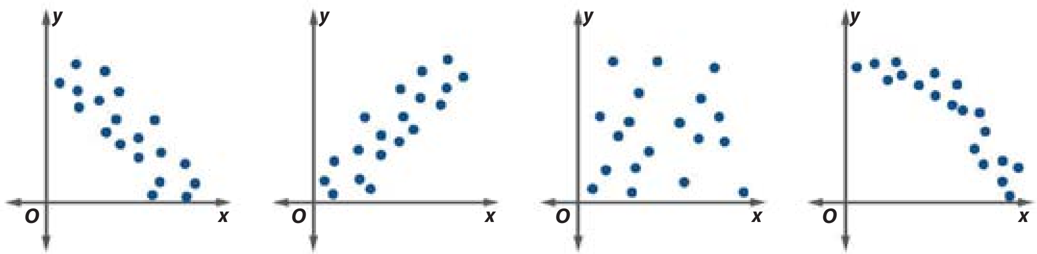
NewVocabulary

correlation
bivariate
explanatory variable
response variable
correlation coefficient
regression line
line of best fit
residual
least-squares regression line
residual plot
influential
interpolation
extrapolation

1 Correlation

Thus far in this chapter, you have graphed, characterized, and used summary statistics to describe distributions of one-variable data sets. You have used sample statistics of such univariate data to make inferences about a population by developing confidence intervals and performing hypothesis tests. **Correlation** is another area of inferential statistics that involves determining whether a relationship exists between two variables in a set of **bivariate** data.

Bivariate data can be represented as ordered pairs (x, y), where x is the independent or **explanatory variable** and y is the dependent or **response variable**. To determine whether there may be a linear, a nonlinear, or no correlation between the variables, you can use a scatter plot.



Negative Linear Correlation | Positive Linear Correlation | No Correlation | Nonlinear Correlation

We say that the data have a strong linear relationship if the points lie close to a straight line and weak if they are widely scattered about the line, but interpreting correlation using a scatter plot tends to be subjective. A more precise way to determine the type and strength of the linear relationship between two variables is to calculate the **correlation coefficient**. A formula for this measure is given below.

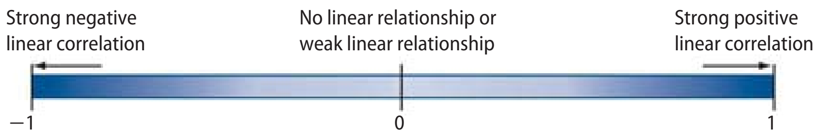
KeyConcept Correlation Coefficient

For n pairs of sample data for the variables x and y, the correlation coefficient r between x and y is given by

$$r = \frac{1}{n-1}\sum\left(\frac{x_i - \bar{x}}{s_x}\right)\left(\frac{y_i - \bar{y}}{s_y}\right),$$

where x_i and y_i represent the values for the ith pair of data, $\bar{x}$ and $\bar{y}$ represent the means of the two variables, and s_x and s_y represent the standard deviations of the two variables.

The correlation coefficient can take on values from -1 to 1. This value indicates the strength and type of linear correlation between x and y as shown in the diagram below.



Notice from the formula that the correlation coefficient is the average of the products of the standardized values for x and the standardized values for y. The correlation coefficient can be tedious to calculate by hand, so we most often rely on computer software or a graphing calculator.

StudyTip

Resistance of Correlation Coefficient Like the mean and standard deviation, r is a nonresistant statistic. It can be affected by outliers.

Example 1 Calculate and Interpret the Correlation Coefficient

SLEEP/GPA STUDY A feature writer for a student newspaper conducts a study to determine whether there is a linear relationship between the average number of hours a student sleeps each night and his or her overall grade point average. The table shows the data that the writer collected. Make a scatter plot of the data, and identify the relationship. Then calculate and interpret the correlation coefficient.

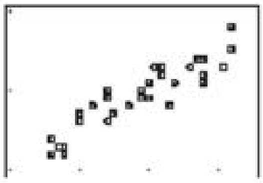
Hours of Sleep	GPA	Hours of Sleep	GPA
6.6	2.2	8.0	2.9
6.6	2.4	8.0	3.1
6.7	2.3	8.1	3.3
6.8	2.3	8.2	3.3
6.8	2.2	8.2	3.2
7.0	2.6	8.3	2.8
7.0	2.7	8.4	3.1
7.2	2.8	8.6	3.3
7.4	2.6	8.7	3.4
7.4	3.0	8.8	3.1
7.4	2.9	8.8	3.2
7.5	2.7	8.8	3.4
7.7	2.8	9.1	3.3
7.9	2.9	9.2	3.8
7.9	3.0	9.2	3.5

Step 1 Graph a scatter plot of the data.

Enter the data into L1 and L2 on your calculator. Then turn on Plot1 under the STAT PLOT menu and choose using L1 for the Xlist and L2 for the Ylist. Graph the scatterplot by pressing ZoomStat or by pressing GRAPH and adjusting the window manually (Figure 11.7.1). From the graph, it appears that the data have a positive linear correlation.

Step 2 Calculate and interpret the correlation coefficient.

Press STAT and select LinReg(ax+b) under the CALC menu (Figure 11.7.2). The correlation coefficient r is about 0.9148. Because r is close to 1, this suggests that the data may have a strong positive linear correlation. This numerical assessment of the data is consistent with our graphical assessment.



[6, 9.5] scl: 1 by [2, 4] scl: 1

Figure 11.7.1

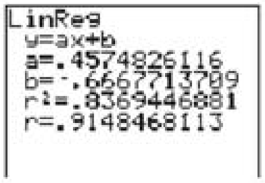


Figure 11.7.2

GuidedPractice

1. **METEOROLOGY** A weather program is featuring a special on a city where a study was conducted to determine whether there is a linear relationship between the average monthly rainfall and temperature. The table in Figure 11.7.3 shows the data collected. Make a scatter plot of the data. Then calculate and interpret the correlation coefficient for the data.

Rain (in.)	Temp. (°F)
5.35	41.3
4.03	44.3
3.77	46.6
2.51	50.4
1.84	56.1
1.59	61.4
0.85	65.3
1.22	65.7
1.94	60.8
3.25	53.5
5.65	46.3
6.00	41.6

Figure 11.7.3

In Example 1, the data collected represent just a sample of the entire school population; therefore, r represents a *sample correlation coefficient*. In order for r to be a valid estimate of the *population correlation coefficient* ρ, the following assumptions must be valid.

- The variables x and y are *linearly* related.
- The variables are *random* variables.
- The two variables have a *bivariate normal distribution*. That is, x and y each come from a normally distributed population.

We would like to use the value of r to make an inference about the relationship between the variables x and y for the entire population. In order to do that, we need to determine whether the value of $|r|$ is great enough to conclude that there is a significant relationship between x and y.

ReadingMath

Population Correlation Coefficient The Greek letter ρ used to represent the population correlation coefficient is pronounced *rho*.

To make this determination, you can perform a hypothesis test. The null and alternative hypotheses for a two-tailed test of the population correlation coefficient ρ are as follows.

H_0: $\rho = 0$ — There is no correlation between the x and y variables in the population.

H_a: $\rho \neq 0$ — There is a correlation between the x and y variables in the population.

We can use a t-test as described below to test the significance of the correlation coefficient.

KeyConcept Formula for the t-Test for the Correlation Coefficient

For a t-test of the correlation between two variables, the test statistic for ρ is the sample correlation coefficient r and the standardized test statistic t is given by

$$t = r\sqrt{\frac{n-2}{1-r^2}}, \text{ where } n - 2 \text{ is the degrees of freedom.}$$

Real-WorldLink

About 40 million people in the U.S. suffer from sleep problems every year. Lack of sleep can have many serious medical consequences.

Source: The National Center on Sleep Disorders Research

Real-World Example 2 Test for Significance

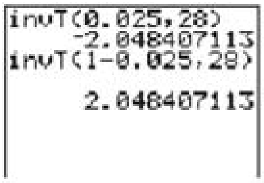
SLEEP/GPA STUDY In Example 1, you calculated the correlation coefficient r for the 30 pairs of student sleep and GPA data to be about 0.9148. Test the significance of this correlation coefficient at the 5% level.

Step 1 State the hypotheses.

H_0: $\rho = 0$ $\qquad$ H_a: $\rho \neq 0$

Step 2 Determine the critical values.

Testing for significance at the 5% level means that $\alpha = 0.05$. Since this is a two-tailed test, the critical values are determined by $\frac{\alpha}{2}$ or 0.025. Using a graphing calculator, the critical values for $\alpha = 0.025$ with $30 - 2$ or 28 degrees of freedom are $t = \pm 2.0$.



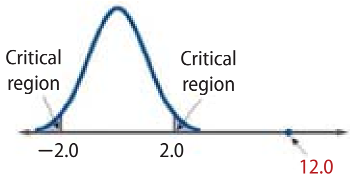
Step 3 Calculate the test statistic.

$$t = r\sqrt{\frac{n-2}{1-r^2}} \qquad \text{Calculate the test statistic for } \rho.$$

$$= 0.9148\sqrt{\frac{30-2}{1-(0.9148)^2}} \text{ or about } 12.0 \qquad r = 0.9148 \text{ and } n = 30$$

Step 4 Reject or fail to reject the null hypothesis.

Since $12.0 > 2.0$, the statistic falls within the critical region and the null hypothesis is rejected.



At the 5% level, there is enough evidence to conclude that there is a significant correlation between the average amount of sleep a student gets each night and his or her overall GPA.

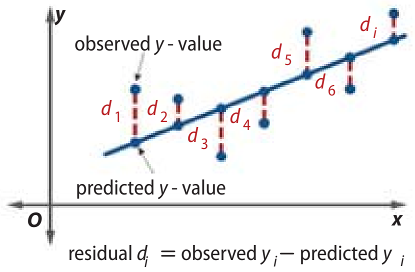
GuidedPractice

2. **METEOROLOGY** In the Guided Practice for Example 1, you calculated the correlation coefficient r for the 12 pairs of rainfall and temperature data. Test the significance of this correlation coefficient at the 10% level.

2 Linear Regression

Once the correlation between two variables is determined to be significant, the next step is to determine the equation of the **regression line**, also called a **line of best fit**. The regression line describes how the response variable y changes as the explanatory variable x changes.

While many lines of best fit can be drawn through a set of points, the one used most often is determined by specific criteria. Consider the scatter plot and regression line shown. The difference d between an observed y-value and its predicted y-value on the regression line is called a **residual**.



Residuals are positive when the observed value is above the line, negative when the observed value is below the line, and zero when it is on the line. The **least-squares regression line** is the line for which the sum of the squares of these residuals is a minimum.

ReadingMath

Regression Equation Notation The symbol $\hat{y}$ is read *y hat* and is used to emphasize that the equation gives the predicted and not the actual response y for any x.

KeyConcept Equation of the Least-Squares Regression Line

The equation of the least-squares regression line for an explanatory variable x and response variable y is $\hat{y} = ax + b$.

The slope a and y-intercept b in this equation are given by

$$a = r\frac{s_y}{s_x} \text{ and } b = \bar{y} - a\bar{x},$$

where r represents the correlation coefficient between the two variables, $\bar{x}$ and $\bar{y}$ represent their means, and s_x and s_y represent their standard deviations.

As with the correlation coefficient, it is not necessary to calculate the least-squares regression equation by hand. Computer software or a graphing calculator will provide the slope a and the y-intercept b of the least-squares regression line for keyed-in values of the variables.

Example 3 Find the Least-Squares Regression Line

SLEEP/GPA STUDY Find the equation of the regression line for the data used in Example 1. Interpret the slope and intercept in context. Then assess the fit of the modeling equation by graphing it, along with the scatter plot of the data, in the same window.

Using the same screen you used to obtain the correlation coefficient (Figure 11.7.4), the least-squares regression equation is approximately $\hat{y} = 0.457x - 0.667$. The slope $a = 0.457$ indicates that for each additional hour of sleep, a student will raise his or her GPA by 0.457 point. The y-intercept $b = -0.667$ indicates that when a student averages no sleep, his or her GPA will be less than 0, which is not possible.

Since the data appear to be randomly scattered about the line $\hat{y} = 0.457x - 0.667$, this regression line appears to be a good fit for the data (Figure 11.7.5).

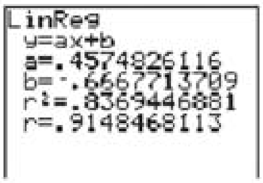
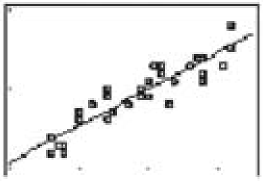


Figure 11.7.4



[6, 9.5] scl: 1 by [2, 4] scl: 1

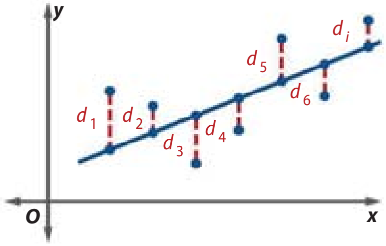
Figure 11.7.5

GuidedPractice

3. **METEOROLOGY** Find the equation of the regression line for the rainfall and temperature data used in the Guided Practice for Example 1. Interpret the slope and intercept in context. Then assess the fit of the modeling equation by graphing it and the scatter plot of the data in the same window.

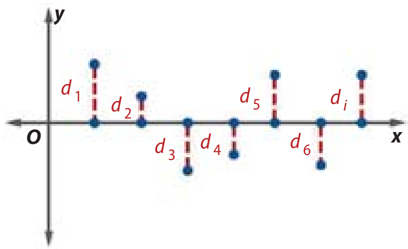
A least-squares regression line describes the overall pattern in a set of bivariate data. As with univariate data analysis, you should always be on the lookout for striking deviations, or outliers, from this pattern. Remember that the residuals measure how much the data deviate from the regression line.

Scatterplot with Regression Line



Examining a scatter plot of the residuals, called a **residual plot**, can help you assess how well the regression line describes the data. In a residual plot, the horizontal line at zero corresponds to the regression line. You can create a residual plot using your graphing calculator. If the plot of the residuals appears to be randomly scattered and centered about $y = 0$, the use of a linear model for the data is supported. If the plot displays a curved pattern, the use of a linear model would not be supported.

Residual Plot



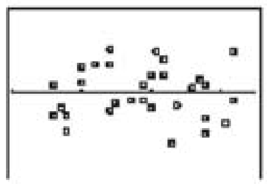
StudyTip

Residuals While residuals can be calculated from any regression line fitted to the data, the residuals from the least-squares regression line have a special property. The mean of the least-squares residuals will always be zero.

Example 4 Graph and Analyze a Residual Plot

SLEEP/GPA STUDY Graph and analyze the residual plot for the average sleep hours and GPA data in Example 1 to determine whether the linear model found in Example 3 is appropriate.

After calculating the least-squares regression line in Example 3, you can obtain the residual plot of the data by turning on Plot2 under the STAT PLOT menu and choosing ⊡, using L1 for the Xlist and RESID for the Ylist. You can obtain RESID by pressing [2nd] [STAT] and selecting RESID from the list of names. Graph the scatter plot of the residuals by pressing ZoomStat.



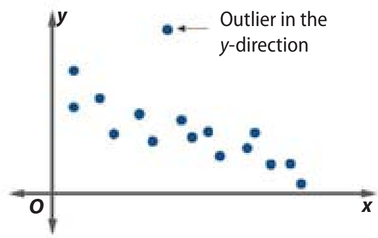
[6, 9.5] scl: 1 by [−0.5, 0.5] scl: 1

The residuals appear to be randomly scattered and centered about the regression line at $y = 0$. This supports the claim that the use of a linear model is appropriate.

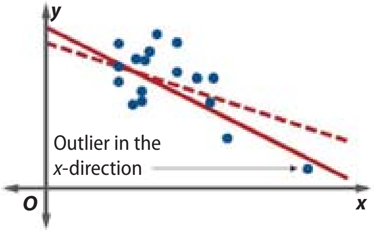
GuidedPractice

4. **METEOROLOGY** Graph and analyze the residual plot for the rainfall and temperature data to determine if the linear model found in the Guided Practice for Example 3 is appropriate.

The residual plot magnifies deviations of the data points from the regression line, making it easier to see outliers in the data that lie in the y-direction. Outliers in the y-direction can indicate errors in data recording or unique cases, especially when describing societal trends or behavioral traits.



Outliers in the x-direction can have a strong influence on the position of a regression line. In the figure, two least-squares regression lines are shown. The solid line is calculated using all the data, while the dashed line is calculated leaving out the outlier in the x-direction. Notice that leaving out this point noticeably moves the regression line.



StudyTip

Influence The influence of an outlier is not a yes or no question. It is a matter of degree and is therefore subjective.

An individual data point that substantially changes a regression line is said to be **influential**. Outliers in the x-direction are often influential to the least-squares regression line. To determine if a point is an influential outlier, calculate and graph regression lines with and without this point. The point is influential if there is a substantial difference in the positions of the regression lines when the point is removed.

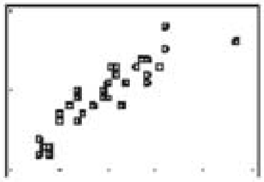
Example 5 Identify an Influential Outlier

SLEEP/GPA STUDY Suppose the feature writer in Example 1 conducting the sleep/GPA study later received the additional piece of data listed in the table, which is an outlier.

Hours of Sleep	GPA
10.7	3.6

a. Make a new scatter plot of the sleep/GPA data that includes the additional data point.

Add the data point to the end of L1 and L2 and then graph the data, adjusting your window as necessary. From the graph you can see that this point is an outlier in the x-direction.



[6, 11] scl: 1 by [2, 4] scl: 1

b. Calculate the correlation coefficient and least squares regression line with this outlier. Describe the effect this outlier has on the strength of the correlation and on the slope and intercept of the regression line.

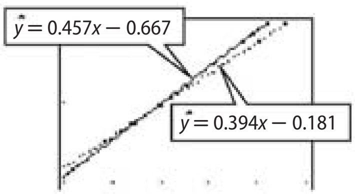
Original data: $r \approx 0.9148$ $\hat{y} = 0.457x - 0.667$

Data with outlier: $r \approx 0.8934$ $\hat{y} = 0.394x - 0.181$

The outlier has reduced the strength of the correlation. The change in the slope of the regression equation has caused the rate at which a student's GPA is raised due to additional sleep to drop from 0.457 points per hour to 0.394 per hour. At the same time, this outlier has raised the y-intercept, indicating that a student who gets no sleep will have a GPA close to 0.

c. Plot both regression lines in the same window. Then state whether the outlier is influential. Explain your reasoning.

The graph of the regression lines shows that the regression line moves more than a small amount when the outlier is added. Therefore, the outlier (10.7, 3.6) is influential.



[6, 11] scl: 1 by [2, 4] scl: 1

GuidedPractice

5. **METEOROLOGY** Suppose the value (2.51, 50.4) for the rainfall and temperature data from Guided Practice 1 was replaced with (0.5, 50.4).
   - **A.** Make a scatter plot of the original temperature/rainfall data that includes this outlier.
   - **B.** Calculate the correlation coefficient and least squares regression line with this outlier. Describe the effect this outlier has on the strength of the correlation and on the slope and intercept of the regression line.
   - **C.** Plot both regression lines in the same window. Then state whether the outlier is influential. Explain your reasoning.

WatchOut!

Making Predictions Do not use a least-squares regression line to make predictions unless a linear model is appropriate and the correlation coefficient is significant. Otherwise, these predictions would be meaningless.

Once you determine that the linear correlation coefficient for a set of data is significant and you find the least-squares regression line, you can then use the equation to make predictions over the range of the data. Making such predictions is called **interpolation**. Using the equation to make predictions far outside the range of the x-values you used to obtain the regression line is called **extrapolation**. Extrapolation should be avoided, since few real-world relationships are linear for all values of the explanatory variable.

Example 6 Predictions with Regression

SLEEP/GPA STUDY The regression equation for the average hours of sleep x and GPA y from Example 3 was $\hat{y} = 0.457x - 0.667$. Use this equation to predict the expected GPA (to the nearest tenth) for a student who averages the following hours of sleep and state whether this prediction is reasonable. Explain.

a. 8 hours

Evaluate the regression equation for $x = 8$ to calculate $\hat{y}$.

$\hat{y} = 0.457x - 0.667$ — Regression equation

$= 0.457(8) - 0.667$ — $x = 8$

$= 3.656 - 0.667$ — Multiply.

$= 2.989$ — Subtract.

Using this model, we would expect that a student averaging 8 hours of sleep would have GPA of about 3.0. This GPA is reasonable since 8 is an x-value in the range of the original data.

b. 10.5 hours

$\hat{y} = 0.457x - 0.667$ — Regression equation

$= 0.457(10.5) - 0.667$ — $x = 10.5$

$= 4.7985 - 0.667$ — Multiply.

$= 4.1315$ — Subtract.

Using this model, we would expect that a student averaging 10.5 hours of sleep would have a GPA of about 4.1. This value is not reasonable, since we are extrapolating a y-value for an x-value that falls far outside the range of the original data. It is also not meaningful, since a student cannot earn a GPA higher than a 4.0 in this model.

GuidedPractice

6. **METEOROLOGY** Use the regression equation for the rainfall and temperature data from Guided Practice 3 to predict the expected temperature (to the nearest tenth of a degree) for months with each average rainfall. State whether this prediction is reasonable. Explain.

A. 3 in. **B.** 8 in.

WatchOut!

Correlation vs. Causation Just because two variables are strongly correlated does not necessarily imply a cause-and-effect relationship. A significant correlation indicates only that y is in some way associated with x.

When analyzing bivariate data, follow the steps summarized below.

ConceptSummary Analyzing Bivariate Data

Step 1 Make a scatter plot, and decide whether the variables appear to be linearly related.

Step 2 If they appear to be linearly related, calculate the strength of the relationship by calculating the correlation coefficient.

Step 3 Use a t-test to determine if the correlation is significant.

Step 4 If significant, find the least-squares regression equation that models the data.

Exercises

= Step-by-Step Solutions begin on page R29.

For Exercises 1–6, analyze the bivariate data. (Examples 1-6)

a. **Make a scatter plot of the data, and identify the relationship. Then calculate and interpret the correlation coefficient.**

b. **Determine if the correlation coefficient is significant at the 1%, 5%, and 10% levels. Explain your reasoning.**

c. **If the correlation is significant at the 10% level, state the least-squares regression equation and interpret the slope and intercept in context.**

d. **Graph and analyze the residual plot.**

e. **Identify any influential outliers. Describe the effect the outlier has on the strength of the original correlation and on the slope and intercept of the original regression line.**

f. **If any data were removed, reassess the significance of the correlation at the 10% level and, if still appropriate, recalculate the regression equation.**

g. **Use the regression equation to make the specified predictions. Interpret your results, and state whether the prediction is reasonable. Explain your reasoning.**

1. **FAT GRAMS AND PROTEIN** An athlete wondered if there is a significant linear correlation between grams of fat and grams of protein in various foods. If appropriate, use the data below to predict the amount of protein (per serving) of an item with 1, 5, or 13 grams of fat.

Fat (g)	Protein (g)	Fat (g)	Protein (g)
12	14	9	13
57	30	18	24
9	15	30	25
20	25	18	25
12	15	32	24
39	28		

2. **FIBER AND CALORIES** The following data shows the caloric counts and amount of fiber in a variety of breakfast cereals. Use the data to predict the Calories in a serving of cereal that has 4.5, 5.5, or 7 grams of fiber.

Fiber (g)	Calories	Fiber (g)	Calories
1.5	133.5	1	149
0.5	115.5	1.5	114.5
1	143	0.5	85.5
2.5	109.5	1	116
0	119	1.5	110
0.5	113.5	0	53.5
0.5	102	8	196.5
0.5	117.5	0.5	99.5
6	186.5	6.5	114.5
1	154	3.5	140.5
11	389	0.5	122.5
4	114.5	2	110

3. **EDUCATION AND HEALTH CARE** The following data lists the performance rankings of education and health care in 14 states. If appropriate, use the data to predict the health care ranking if the education ranking is 15, 28, or 42.

Education	Health Care	Education	Health Care
1	45	8	35
2	48	9	18
3	50	10	13
4	37	11	20
5	39	12	28
6	26	13	15
7	21	14	29

4. **WEIGHT AND MPG** A shopper wants to determine if there is a significant linear correlation between the weight of cars and their highway fuel efficiency. If appropriate, use the data below to predict the gas mileage of automobiles that weigh 2900, 3300, and 4000 pounds.

Weight (lb)	MPG	Weight (lb)	MPG
3450	32	3460	28
3216	32	2897	36
2636	34	2805	32
2690	40	3067	28
2875	51	2716	31
2403	36	2595	38
2972	35	2326	39
2811	34	2911	30

5. **GRADUATION AND UNEMPLOYMENT** An economist took a sample of the graduation rates and unemployment rates of various states in a given year. If appropriate, use the data below to predict the unemployment rate if the graduation rates are 70%, 80%, or 90%.

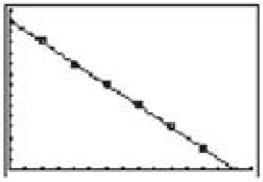
Graduated	73	85	64	79	68	82
Unemployed	6.9	4.1	3.2	2.9	4.3	5.1
Graduated	71	81	76	64	77	82
Unemployed	4.1	5.5	5	6.8	4.8	5.2

6. **POPULATION AND CRIME** The following data lists the performance rankings of population and crime in 14 states. If appropriate, use the data to predict the crime ranking if the population ranking is 15.

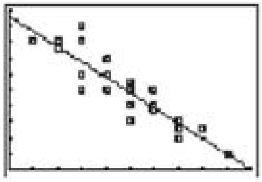
Population	1	2	3	4	5	6	7
Crime	14	15	13	4	5	9	7
Population	8	9	10	11	12	13	14
Crime	11	3	12	10	8	1	6

Match each graph to the corresponding correlation coefficient.

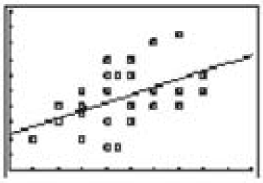
a.



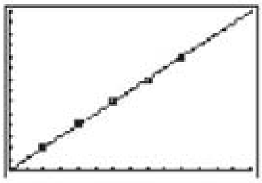
b.



c.



d.



7. $r = -0.90$

8. $r = 0.50$

9. $r = 1.00$

10. $r = -1.00$

11. **INCOME AND DINING OUT** A restaurant is conducting a study to determine the relationship between a person's monthly income and the number of times that person dines out each month.

Income	500	1125	300	750	1250	950
Meals	4	10	3	6	12	8

a. Make a scatter plot of the data, and linearize the data by finding $(x, \ln y)$.

b. Make a scatter plot of the linearized data, and calculate and interpret r.

c. Determine if r is significant at the 5% level.

d. If r is significant, find the least-squares regression equation by using the model for the linearized data to find a model for the original data. (*Hint*: You can review this in Lesson 3-5.)

e. If appropriate, use the regression equation to predict the number of times that a person with a monthly income of \$2000 will dine out. Is the prediction reasonable? Explain your reasoning.

12. **ADS AND SALES** An advertising firm wants to determine the strength of the relationship between the number of television ads aired each week and the amount of sales (in thousands of dollars) of the product.

Ads	2	3	5	7	7
Sales (\$)	3	4	6	8	9
Ads	8	9	10	10	12
Sales (\$)	10	12	12	13	15

a. Make a scatter plot of the data, and identify the relationship. Then find the correlation coefficient.

b. Determine if the correlation coefficient is significant at the 10% level. If so, find the least-squares regression equation.

c. Suppose the firm airs 15 ads during one week and 18 ads during the following week, and each ad spot costs \$500. Make a prediction about the increase in profit from the first week to the second week.

13. **MULTIPLE REPRESENTATIONS** In this problem, you will investigate the *coefficient of determination*.

a. **GRAPHICAL** Make a scatter plot of the data below. Then calculate the correlation coefficient r.

x	1	2	3	4	5	6
y	4	9	12	15	20	24

b. **NUMERICAL** Find the mean $\overline{y}$ of the y-values.

c. **NUMERICAL** Determine the least squares regression equation, and find the predicted $\hat{y}$-values by substituting each value of x into the equation.

d. **NUMERICAL** Use the following formulas to find the total variation $\Sigma(y - \overline{y})^2$, explained variation $\Sigma(\hat{y} - \overline{y})^2$ and unexplained variation $\Sigma(y - \hat{y})^2$.

e. **NUMERICAL** The coefficient of determination is given by $r^2 = \dfrac{\text{explained variation}}{\text{total variation}}$. Use the formula and your answers from part **d** to find r^2.

f. **ANALYTICAL** If the explained variation is the variation that can be explained by the relationship between x and y, what do you think the value of the coefficient of determination that you found means?

H.O.T. Problems Use Higher-Order Thinking Skills

REASONING **Determine whether each statement is *true* or *false*. Explain your reasoning.**

14. An r value of -0.85 indicates a stronger linear correlation than an r value of 0.75.

15. If the null hypothesis is rejected, it means that the value of ρ is not significantly different from 0.

16. **CHALLENGE** Consider two sets of bivariate data, C and D, which represent exponential relationships. With an exponential regression, the value of the base b in C is the reciprocal of the value of b in D. The correlation coefficients for each are equal to 0.99. What is the relationship of the linearized regression lines of C and D?

17. **REASONING** Consider the data set below where row A represents the explanatory variable and row B represents the response variable.

A	21	30	44	49	52	59
B	114	127	148	154	169	179

a. Make a scatter plot of the data. Then determine the equation for the least squares regression line and graph it in the same window as the scatter plot.

b. Interchange A and B and repeat part **a**.

c. What effect does switching the explanatory and response variables have on the regression line?

18. **WRITING IN MATH** Describe the strengths and weaknesses of the correlation coefficient as a measure of linear correlation for a set of bivariate data.

Spiral Review

19. FOOTBALL Alexi claims that she can throw a football at least 55 yards. After 37 throws, her average distance is 57.7 yards with a standard deviation of 3.6 yards. Is there enough evidence to reject Alexi's claim at $\alpha = 0.05$? Explain your reasoning. (Lesson 11-6)

20. BOWLING Sonia and Pearl want to compare their bowling scores. They recorded their scores for 16 games as shown. (Lesson 11-5)

a. Calculate the mean and sample standard deviation for each data set.

b. Construct two 99% confidence intervals for the average score for both Sonia and Pearl.

c. Make a statement comparing the effectiveness of the two intervals.

Sonia		Pearl	
112	109	88	169
98	116	129	190
143	131	146	99
109	98	170	108
121	122	95	181
84	128	111	183
106	121	108	122
100	107	181	99

If possible, find the sum of each infinite geometric series. (Lesson 10-3)

21. $a_1 = 4, r = \frac{5}{7}$

22. $a_1 = 14, r = \frac{7}{3}$

23. $16 + 12 + 9 + \ldots$

Write an explicit formula and a recursive formula for finding the *n*th term of each arithmetic sequence. (Lesson 10-2)

24. $10, 26.5, 43, \ldots$

25. $15, -9, -33, \ldots$

26. $3, \frac{11}{3}, \frac{13}{3}, \ldots$

Express each complex number in polar form. (Lesson 9-5)

27. $6 - 8i$

28. $-4 + i$

29. $3 + 2i$

Determine whether each pair of vectors are parallel. (Lesson 8-5)

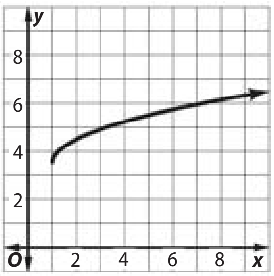
30. $\mathbf{g} = \langle 3, 4, -6 \rangle$, $\mathbf{h} = \langle 9, 12, -18 \rangle$

31. $\mathbf{j} = \langle 9, -15, 11 \rangle$, $\mathbf{k} = \langle -14, 10, 7 \rangle$

32. $\mathbf{n} = \langle -16, -8, -13 \rangle$, $\mathbf{p} = \langle -15, 9, 5 \rangle$

Skills Review for Standardized Tests

33. SAT/ACT Which of the following must be true about the graph?



I. The domain is all real numbers.

II. The function is $y = \sqrt{x} + 3.5$.

III. The range is about $\{y \mid y \geq 3.5\}$.

A I only

B II and III

C I, II, and III

D II only

E III only

34. The table shows the total attendance for minor league baseball in some recent years. Which of the following is a regression equation for the data?

Year	Attendance (millions)
1990	18.4
1995	25.2
2000	33.1
2005	37.6

F $y = 1.31x - 2588.15$

G $y = 1.46x - 2588.15$

H $y = 1.31x - 18.4$

J $y = 1.46x - 18.4$

35. FREE RESPONSE For the following problem, consider a real-life situation that exhibits the characteristics of exponential or logistic growth or decay.

a. Identify the situation and the type of growth or decay that it represents.

b. Pose a question or make a claim about the situation.

c. Make a hypothesis to the answer of the question.

d. Develop, justify, and implement a method to collect, organize, and analyze the related data.

e. Extend the nature of collected, discrete data to that of a continuous function that describes the known data set.

f. Generalize the results and make a conclusion.

g. Compare the hypothesis and the conclusion.

EXTEND

11-7 Graphing Technology Lab
Median-Fit Lines

Objective

- Use a TI-Nspire technology to find a median-fit line to model a relationship shown in a scatter plot.

In previous lessons, you have used regression equations to represent a set of data. Another type of regression used to model data is a *median-fit line*.

A median-fit line is found by dividing a set of data into three equal-sized groups and using the medians of those groups to determine a regression equation for the data.

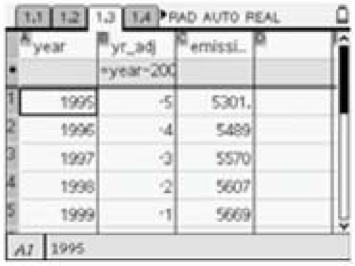
Activity 1 Draw a Median-Fit Line

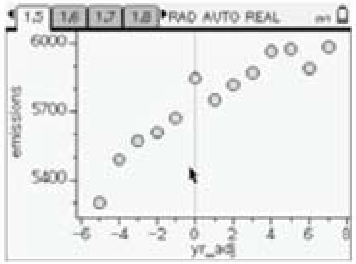
Use the data in the table to draw a median-fit line.

U.S. Energy Related Carbon Dioxide Emissions (million metric tons)			
Year	Emissions	Year	Emissions
1995	5301	2002	5820
1996	5489	2003	5872
1997	5570	2004	5966
1998	5607	2005	5974
1999	5669	2006	5888
2000	5848	2007	5984
2001	5754		

Source: Energy Information Administration

Step 1 Enter the data in a spreadsheet. Then make a scatter plot of the data. Let the x-axis represent the number of years where 0 represents the year 2000 and the y-axis the metric tons of carbon dioxide.



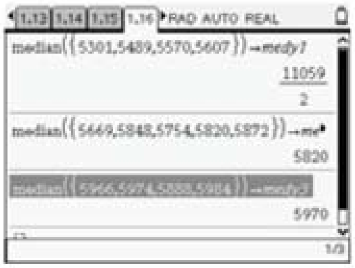


Step 2 Divide the data into three relatively equal and symmetric groups. The second group will have 5 data values, and the other groups will have 4. Then find the medians for the x- and y-values of each group.

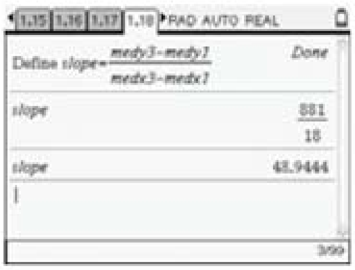
Group 1 median: $(-3.5, 5529.5)$

Group 2 median: $(1, 5820)$

Group 3 median: $(5.5, 5970)$



Step 3 The median-fit line uses the median points from the 1st and 3rd groups to determine slope and the average of the three median points as a point on the line. By using the point-slope form, $y = m(x - a) + b$, where m = slope and (a, b) is the average, you can form the median-fit line.



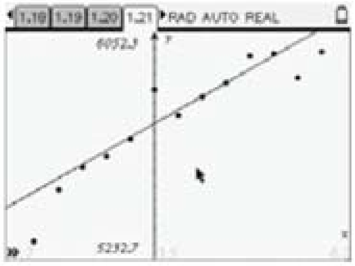
TechnologyTip

Icons Use different icons for the median points to easily distinguish them from the regular data points. Grab each point and select Attributes to change the icon type.

StudyTip

Geometric Interpretation Geometrically, the three median points determine a triangle and the average of these x- and y-values is the centroid of the triangle.

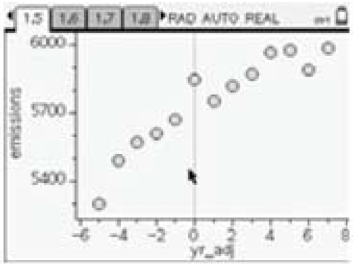
Step 4 Define the average of y as $avey = \dfrac{medy1 + medy2 + medy3}{3}$. Define the median-fit line as $med\_med(x) = 48.944(x - 1) + 5773.17$. Then graph the median-fit line.



Activity 2 Calculate a Median-Fit Line

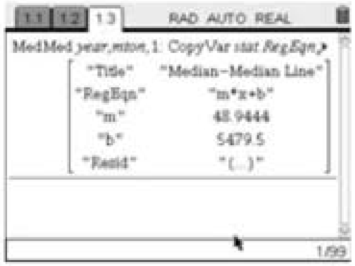
Use the data in Activity 1 to calculate the median-fit line.

Step 1 Remove the three ordered pairs that represent the medians. Then make a scatter plot of the data.

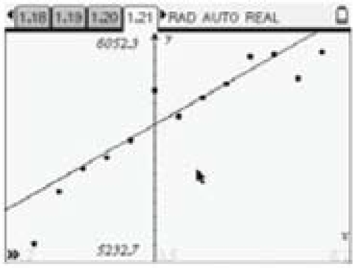


Step 2 Calculate the equation of the median-fit line. Then graph the line.

Open a new Calculator screen. Under the Statistics: Stat Calculations menu, select Median-Median Line. Enter the lists for the x- and y-values.



Notice that the equation of the median-fit line found in Activity 1 is identical to the calculator regression equation.



Analyze the Results

1. Explain the meaning of the slope of median-fit-line in this situation.
2. Is it reasonable to expect this line to represent the data indefinitely? Explain why or why not.
3. How many metric tons of carbon dioxide emissions can be expected in 2015?

CHAPTER 11

Study Guide and Review

Chapter Summary

KeyConcepts

Descriptive Statistics (Lesson 11-1)

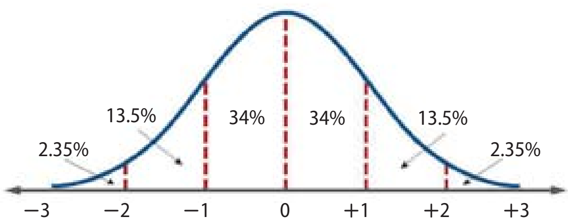
- The three most common shapes for distributions of data are negatively skewed, symmetrical, and positively skewed.

Probability Distributions (Lesson 11-2)

- The probability distribution of a random variable X links each possible value for X with its probability of occurring.

The Normal Distribution (Lesson 11-3)

- The z-value represents the number of standard deviations that a given data value is from the mean, and is given by $z = \frac{X - \mu}{\sigma}$.
- The standard normal distribution is a distribution of z-values with mean 0 and standard deviation 1.



The Central Limit Theorem (Lesson 11-4)

- As the sampling size n increases, the shape of the distribution of the sample means approaches a normal distribution.

Confidence Intervals (Lesson 11-5)

- When $n \geq 30$, $CI = \bar{x} \pm z \cdot \frac{\sigma}{\sqrt{n}}$; when $n < 30$ and σ is unknown, $CI = \sigma \pm t \cdot \frac{s}{\sqrt{n}}$.

Hypothesis Testing (Lesson 11-6)

- The steps to conduct a hypothesis test are as follows.
  - **Step 1** State the hypotheses, and identify the claim.
  - **Step 2** Determine the critical value(s) and region.
  - **Step 3** Calculate the test statistic.
  - **Step 4** Accept or reject the null hypothesis.

Correlation and Linear Regression (Lesson 11-7)

- To analyze bivariate data:
  - **Step 1** Make a scatter plot, and decide whether the variables appear to be linearly related.
  - **Step 2** Calculate the correlation coefficient.
  - **Step 3** Use a t-test to determine if the correlation is significant.
  - **Step 4** Find the least-squares regression equation.

KeyVocabulary

alternative hypothesis (p. 705)
binomial distribution (p. 669)
confidence interval (p. 697)
continuous random variable (p. 664)
correlation (p. 713)
correlation coefficient (p. 713)
critical values (p. 697)
discrete random variable (p. 664)
empirical rule (p. 675)
explanatory variable (p. 713)
extrapolation (p. 719)
hypothesis test (p. 705)
inferential statistics (p. 696)
interpolation (p. 719)
least squares regression line (p. 716)
negatively skewed distribution (p. 654)
normal distribution (p. 674)
null hypothesis (p. 705)
percentiles (p. 658)
positively skewed distribution (p. 654)
probability distribution (p. 665)
random variable (p. 664)
regression line (p. 716)
response variable (p. 713)
sampling distribution (p. 685)
sampling error (p. 686)
standard normal distribution (p. 676)
symmetrical distribution (p. 654)
t-distribution (p. 699)
z-value (p. 676)

VocabularyCheck

Identify the word or phrase that best completes each sentence.

1. The mean is less than the median and the majority of the data are on the right in a (negatively skewed, positively skewed) distribution.
2. A (continuous, discrete) random variable can take on an infinite number of possible values within a specified interval.
3. A distribution of z-values with a mean of 0 and a standard deviation of 1 is called a (binomial, standard normal) distribution.
4. The standard deviation of the sample means is called the (sampling error, standard error of the mean).
5. The (Central Limit Theorem, empirical rule) states that as n increases, the shape of the distribution of the sample means will approach a normal distribution.
6. A single value estimate of an unknown population parameter is called a(n) (point, interval) estimate.
7. The (alternative, null) hypothesis states that there is not a significant difference between a sample value and a population parameter.
8. Using an equation to make predictions far outside the range of the x-values you used to obtain the regression line is called (extrapolation, interpolation).

CHAPTER 11 Study Guide and Review *Continued*

Lesson-by-Lesson Review

11-1 Descriptive Statistics (pp. 654–663)

9. SAT SCORES The table gives the math SAT scores for 24 precalculus students.

SAT Math Scores					
373	437	477	491	503	516
392	454	479	491	508	519
405	463	485	498	508	522
417	470	485	499	513	533

a. Construct a histogram, and use it to describe the shape of the distribution.

b. Summarize the center and spread of the data using either the mean and standard deviation or the five-number summary.

10. RADON GAS The table shows the amount in picocuries per liter of radon gas in a sample of homes.

Amount of Radon (pCi/L)				
0.5	1.1	1.9	2.4	4.0
0.7	1.4	2.2	2.5	4.2
1.0	1.5	2.2	2.9	5.4
1.0	1.7	2.2	2.9	6.3
1.1	1.8	2.3	3.1	7.0

a. Construct a box plot, and use it to describe the shape of the distribution.

b. Summarize the center and spread of the data using either the mean and standard deviation or the five-number summary. Justify your choice.

11. MARATHON The table gives the frequency distribution of the completion times for the Boston Marathon for the first 322 women finishers. Construct a percentile graph. Estimate the percentile rank for those finishing below 3 hours, and interpret its meaning.

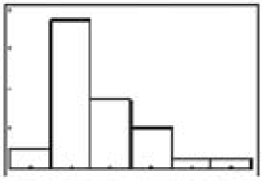
Time (hours)	Runners
2:45–2:49:59	3
2:50–2:54:59	4
2:55–2:59:59	28
3:00–3:04:59	35
3:05–3:09:59	54
3:10–3:14:59	80
3:15+	118

Example 1

BACKPACKS The table shows the weight of school backpacks for a sample of high school students.

Average Backpack Weight (lb)					
11.5	15.0	16.0	17.0	19.0	24.5
12.5	15.5	16.0	17.5	21.0	25.0
14.5	15.5	16.5	18.0	21.0	25.0
14.5	15.5	17.0	18.0	21.5	27.0
15.0	16.0	17.0	18.5	23.5	30.0

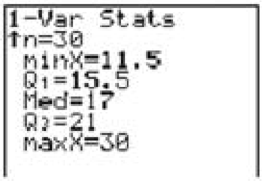
a. Construct a histogram, and use it to describe the shape of the distribution.



[10, 34] scl: 4 by [0, 16] scl: 4

The graph is positively skewed. Most of the backpacks appear to weigh between 14 and 22 pounds, with a few that are heavier, so the tail of the distribution trails off to the right.

b. Summarize the center and spread of the data using either the mean and standard deviation or the five-number summary.



The distribution of data is skewed; therefore, the five-number summary can be used to describe the distribution. The five-number summary indicates that while the weights range from 11.5 to 30 pounds, the median weight is 17 pounds and half of the weights are between 15.5 and 21 pounds.

11-2 Probability Distributions (pp. 664–673)

Classify each random variable X as *discrete* or *continuous*. Explain your reasoning.

12. X represents the number of people attending an opening show of a new movie on a given day.

13. X represents the amount of blood donated per person at a recent blood drive.

14. **FAMOUS PEOPLE** In a survey, 63% of adults said they recognized a certain famous athlete. Five adults are selected at random and asked if they recognize the athlete.

a. Construct and graph a binomial distribution for the random variable X representing the number of adults who recognized the athlete.

b. Find the probability that more than 2 adults recognized the athlete.

15. **DOGS** Find the variance and standard deviation of the probability distribution for the number of dogs per household in Greenville, South Carolina.

Dogs	Frequency
0	17,519
1	2720
2	1614
3	774
4	333

Example 2

GRAPHING In a school survey, 45% of the students said that they knew how to graph a conic. Five students chosen at random are asked if they can graph a conic.

a. Construct and graph a binomial distribution for the random variable X representing the number of students who said they could graph a conic.

Here $n = 5$, $p = 0.45$, and $q = 1 - 0.45$ or 0.55.

$P(0) = {}_5C_0 \cdot 0.45^0 \cdot 0.55^5 \approx 0.050$

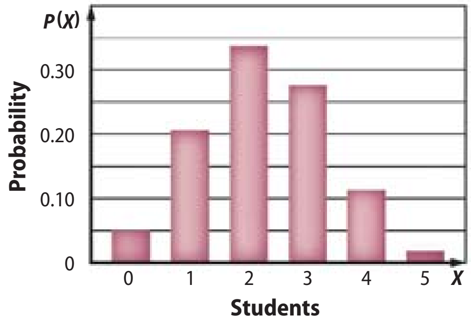
$P(1) = {}_5C_1 \cdot 0.45^1 \cdot 0.55^4 \approx 0.206$

$P(2) = {}_5C_2 \cdot 0.45^2 \cdot 0.55^3 \approx 0.337$

$P(3) = {}_5C_3 \cdot 0.45^3 \cdot 0.55^2 \approx 0.276$

$P(4) = {}_5C_4 \cdot 0.45^4 \cdot 0.55^1 \approx 0.113$

$P(5) = {}_5C_5 \cdot 0.45^5 \cdot 0.55^0 \approx 0.018$



b. Find the probability that fewer than three of the students interviewed could graph a conic.

$P(X < 3) = P(0) + P(1) + P(2)$

$= 0.05 + 0.21 + 0.34$ or 0.60 or 60%

11-3 The Normal Distribution (pp. 674–683)

Find each of the following.

16. z if $X = 1.5$, $\mu = 1.1$, and $\sigma = 0.3$

17. X if $z = 2.34$, $\mu = 105$, and $\sigma = 18$

18. z if $X = 125$, $\mu = 100$, and $\sigma = 15$

19. X if $z = -1.12$, $\mu = 35$, and $\sigma = 3.4$

Find the interval of z-values associated with each area.

20. outside 55%

21. middle 24%

22. middle 96%

23. outside 49%

Example 3

Find z if $X = 36$, $\mu = 31$, and $\sigma = 1.3$.

$z = \dfrac{X - \mu}{\sigma}$ Formula for z-values

$= \dfrac{36 - 31}{1.3}$ $X = 36$, $\mu = 31$, and $\sigma = 1.3$

≈ 3.85 Simplify.

CHAPTER 11 Study Guide and Review *Continued*

11-4 The Central Limit Theorem (pp. 685–694)

24. GRADES The average grade-point average or GPA in a particular school is 2.88 with a standard deviation of approximately 0.67. Find each probability for a random sample of 50 students from that school.

a. the probability that the mean GPA will be less than 2.75

b. the probability that the mean GPA will be greater than 3.05

c. the probability that the mean GPA will be greater than 3.0 but less than 3.75

25. PHOTOGRAPHY A local photographer reported that 55% of seniors had their senior photos taken outdoors. If 15 seniors are selected at random, find the probability that fewer than 5 of the seniors will get their pictures taken outdoors.

Find each of the following if z is the z-value, $\bar{x}$ is the sample mean, μ is the mean of the population, n is the sample size, and σ is the standard deviation.

26. z if $\bar{x} = 5.8$, $\mu = 5.5$, $n = 18$, and $\sigma = 0.2$

27. μ if $\bar{x} = 14.8$, $z = 4.49$, $n = 14$, and $\sigma = 1.5$

28. n if $z = 1.5$, $\bar{x} = 227$, $\mu = 224$, and $\sigma = 10$

29. σ if $z = -2.67$, $\bar{x} = 38.2$, $\mu = 40$, and $n = 16$

Example 4

WEATHER The average annual snowfall for Albany, New York, is 62 inches with a standard deviation of approximately 20 inches. Find the probability that the mean snowfall will be between 60 and 70 inches using a random sample of data for 7 years.

z-value for $\bar{x} = 60$:

$= \frac{60 - 62}{7.56}$ $\bar{x} = 60$, $\mu = 62$, and $\sigma_{\bar{x}} = \frac{20}{\sqrt{7}} \approx 7.56$

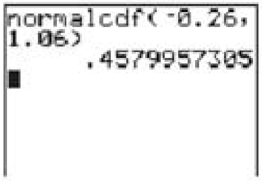
≈ -0.26 Simplify.

z-value for $\bar{x} = 70$:

$= \frac{70 - 62}{7.56}$ $\bar{x} = 70$, $\mu = 62$, and $\sigma_{\bar{x}} = \frac{20}{\sqrt{7}} \approx 7.56$

≈ 1.06 Simplify.

There is a 45.8% probability that the snowfall will be between 60 and 70 inches.



11-5 Confidence Intervals (pp. 696–704)

Determine whether the normal distribution or t-distribution should be used for each question. Then find each confidence interval given the following information.

30. $c = 90\%$, $\bar{x} = 73$, $s = 4.8$, $n = 12$

31. $c = 96\%$, $\bar{x} = 34$, $\sigma = 2.3$, $n = 38$

32. $c = 99\%$, $\bar{x} = 16$, $s = 1.6$, $n = 55$

33. $c = 90\%$, $\bar{x} = 5.8$, $\sigma = 1.1$, $n = 47$

Determine the minimum sample size needed to conduct an experiment that has the given requirements.

34. $c = 90\%$, $\sigma = 3.9$, $E = 0.8$

35. $c = 95\%$, $\sigma = 1.6$, $E = 0.6$

36. $c = 98\%$, $\sigma = 6.8$, $E = 1.2$

37. $c = 92\%$, $\sigma = 10.2$, $E = 3.5$

Example 5

Determine whether the normal distribution or t-distribution should be used to find a 95% confidence interval in which $\bar{x} = 12.8$, $s = 3.8$, and $n = 50$. Then find the confidence interval.

Since $n \geq 30$, the normal distribution should be used.

In a 95% confidence interval, 2.5% of the area lies in each tail. Use a graphing calculator to find z.

$CI = \bar{x} \pm z \cdot \frac{s}{\sqrt{n}}$ Confidence Interval for the Mean

$= 12.8 \pm 1.96 \cdot \frac{3.8}{\sqrt{50}}$ $\bar{x} = 12.8$, $z = 1.96$, $s = 3.8$, and $n = 50$

$= 12.8 \pm 1.05$ Simplify.

The 95% confidence interval is $11.75 < \mu < 13.85$.

11-6 Hypothesis Testing (pp. 705–712)

For each statement, write the null and alternative hypotheses, and state which hypothesis represents the claim.

38. Jenna claims that she did not drive above 50 miles per hour during the entire trip.

39. Rafael claims that he can type faster than 60 words per minute.

40. Natasha claims that on average, it takes her less than 3 days to read a short novel.

41. Grant claims that he can bake at least 6 dozen cookies per hour.

For each claim k, use the specified information to calculate the test statistic and determine whether there is enough evidence to reject the null hypothesis. Then make a statement regarding the original claim.

42. k: $\mu \leq 26.5$, $\alpha = 0.10$, $\bar{x} = 27.8$, $s = 1.0$, $n = 46$

43. k: $\mu = 56$, $\alpha = 0.05$, $\bar{x} = 58.9$, $s = 6.7$, $n = 98$

44. k: $\mu < 18$, $\alpha = 0.01$, $\bar{x} = 17.6$, $s = 0.8$, $n = 26$

45. k: $\mu \geq 39$, $\alpha = 0.10$, $\bar{x} = 38.6$, $s = 2.6$, $n = 42$

Example 6

For claim k, use the specified information to calculate the test statistic and determine whether there is enough evidence to reject the null hypothesis. Then make a statement regarding the original claim.

k: $\mu \geq 62$, $\alpha = 0.05$, $\bar{x} = 61.5$, $s = 4.3$, $n = 70$

State the null and alternative hypotheses, and identify the claim.

H_0: $\mu \geq 62$ (claim) H_α: $\mu < 62$

Determine the critical value(s) and region.

Use the z-value since $n \geq 30$ and a left-tailed test since $\mu < 62$. Since $\alpha = 0.05$, you can use a graphing calculator to find $z = -1.645$.

Calculate the test statistic.

$z = \frac{61.5 - 62}{0.51} \approx -0.98$ $\bar{x} = 61.5$, $\mu = 62$, and $\sigma_{\bar{x}} = \frac{4.3}{\sqrt{70}}$

Reject or fail to reject the null hypothesis.

H_0 is not rejected since the test statistic does not fall within the critical region. Therefore, there is not enough evidence to reject the claim.

11-7 Correlation and Linear Regression (pp. 713–722)

46. GRADES The table shows the pre-test and final grades for a high school college prep class. (x = pre-test, y = final)

Scores for a College Prep Class

x	y	x	y	x	y	x	y
86	3.5	77	2.5	85	3.0	62	1.9
70	3.0	97	3.9	85	3.8	92	3.6
100	4.0	79	3.0	68	2.2	84	3.0
87	3.8	69	2.4	73	2.4	84	3.6
99	4.0	67	2.1	91	3.7	74	2.8

a. Make a scatter plot of the data, and identify the relationship. Then calculate and interpret the correlation coefficient.

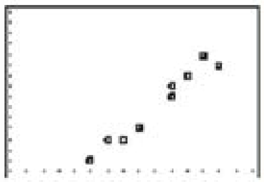
b. Test the significance of this correlation coefficient at the 10% level.

Example 7

HEIGHT **The table shows the heights of brothers and sisters. Make a scatter plot of the data and identify the relationship. Then calculate and interpret the correlation coefficient.**

Brother	71	68	66	67	70
Sister	69	64	63	63	68
Brother	71	70	73	72	65
Sister	69	67	70	71	61

The correlation coefficient r is about 0.9773. Since r is close to 1, this suggests that the data have a strong positive linear correlation. This numerical assessment of the data agrees with our graphical assessment.



[60, 75] scl: 1 by [60, 75] scl: 1

Applications and Problem Solving

47. **SPORTS** The body-fat levels of 20 professional basketball players are shown. (Lesson 11-1)

Body-Fat Levels (%)			
3.4	5.5	6.1	4.8
8.3	7.7	6.5	6.5
4.9	3.7	3.9	4.0
7.3	8.9	9.5	9.8
3.9	7.1	6.3	6.1

a. Construct a histogram, and use it to describe the shape of the distribution.

b. Summarize the center and spread of the data using either the mean and standard deviation or the five-number summary. Justify your choice.

48. **EXERCISE** The number of hours that a sample of students exercises each week is shown. (Lesson 11-1)

Time Spent Exercising (hours)		
3	2.5	0
1.5	3	2
3.5	2	0
1.5	9.5	0
8	0.5	1.5
1	10	4

a. Construct a box plot, and use it to describe the shape of the distribution.

b. Summarize the center and spread of the data using either the mean and standard deviation or the five-number summary. Justify your choice.

49. **AP CLASSES** The table shows the number of AP classes per senior. Find the mean, variance, and standard deviation of this distribution. (Lesson 11-2)

X	0	1	2	3	4
Frequency	12	18	25	19	11

50. **IQ** IQs for a group of people are normally distributed with a mean of 105 and a standard deviation of 22. Find the probability that a randomly chosen person will have an IQ that corresponds to each of the following. (Lesson 11-3)

a. above 101

b. below 94

c. between 110 and 120

51. **COOKIES** The number of chocolate chips in a cookie is normally distributed with $\mu = 25$ and $\sigma = 3$. Find each of the following. (Lesson 11-3)

a. $P(X < 35)$

b. $P(21 < X < 29)$

c. $P(X > 15)$

52. **WRESTLING** The average number of fans attending East High School's wrestling meets is normally distributed with $\mu = 88$ and $\sigma = 16$. If 6 random meets are selected, find the probability that the mean of the sample would be more than 90 fans. (Lesson 11-4)

53. **EXERCISE** A sample of 58 students found that on average, the students spend 185 minutes engaged in physical activity each week. Assume that the standard deviation from a recent study was 28 minutes. Estimate the mean time students spend engaged in physical activity each week using a confidence interval given a 95% confidence level. (Lesson 11-5)

54. **FLIGHT** An airline claims that its flights from Cleveland to Texas take less than 3.0 hours. A random sample of 30 flights found an average time of 2.9 hours and a standard deviation of 0.25 hour. Determine whether the airline's claim is supported at $\alpha = 0.05$. (Lesson 11-6)

a. Write the null and alternative hypotheses, and state which hypothesis represents the claim.

b. Calculate the test statistic.

c. Determine whether there is enough evidence to reject the null hypothesis.

d. Make a statement regarding the original claim.

55. **SCORES** **The table shows the aptitude and writing test scores for a class over the same material.** (Lesson 11-7)

Aptitude					Writing				
135	146	153	154	139	26	33	55	50	32
131	149	137	133	149	25	44	31	31	34
141	164	146	149	147	32	47	37	46	36
152	143	146	141	136	47	36	35	28	28
154	151	155	140	143	36	48	36	33	42
148	149	141	137	135	32	32	29	34	30

a. Make a scatter plot of the data, and identify the relationship. Then calculate and interpret the correlation coefficient.

b. Test the significance of this correlation coefficient at the 5% level.

c. Find the equation of the regression line.

d. Use this equation to predict the writing score for a student who scored a 142 on the aptitude test.

CHAPTER 11 Practice Test

1. **RACING** The ages of the last 20 winners of the Indianapolis 500 are shown.

Age (years)									
24	26	28	33	40	25	27	30	36	42
26	27	32	35	43	26	27	33	38	46

   a. Construct a histogram, and use it to describe the shape of the distribution.

   b. Summarize the center and spread of the data using either the mean and standard deviation or the five-number summary.

2. **TELEVISION** The number of televisions per household for 100 students is shown.

Televisions	0	1	2	3	4	5
Frequency	1	3	21	53	16	6

   a. Use the frequency distribution to construct and graph a probability distribution for the random variable X.

   b. Find the mean score, and interpret its meaning in the context of the problem situation.

   c. Find the variance and standard deviation of the probability distribution.

3. **FOOD** Ms. Martinez's Spanish class took a poll to find how many drive-through trips students made in a week.

X	0	1	2	3	4	5
Frequency	10	16	12	22	8	2

   a. Use the frequency distribution of the results to construct and graph a probability distribution for the random variable X, rounding each probability to the nearest hundredth.

   b. Find the mean of the probability distribution.

   c. Find the variance and standard deviation.

4. **VACATION** In the summer, the average temperature at a Caribbean vacation resort is 89°F with a standard deviation of 4.8°F. For a randomly selected day, find the probability that the temperature will be as follows.

   a. above 72°F

   b. below 68°F

   c. between 85°F and 93°F

PACKAGING The mean weight of a box of cereal is 362 grams and standard deviation of 5. If a random selection of 5 boxes is sampled, find the following.

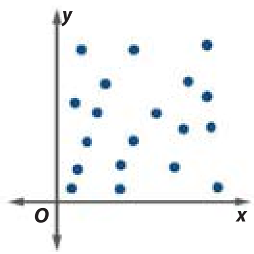
5. probability that the mean weight is less than 355

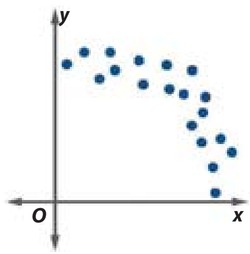
6. probability that the mean weight is greater than 370

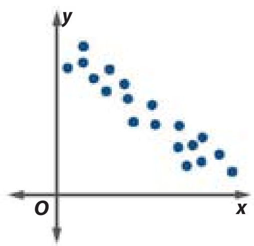
7. **CONCESSIONS** A survey of 97 movie patrons found that customers spent an average of $12.50 at the concessions counter. Assume that the standard deviation from a recent study was $2.25. Estimate the mean amount of money customers spend given a 95% confidence level.

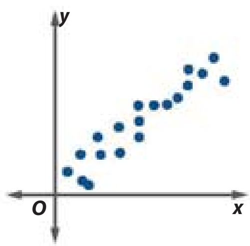
8. **RENT** Phil claims that the average college student spends less than $400 a month on rent. A sample of 48 students found that students spent an average of $385 on rent each month and a standard deviation of $30. At $\alpha = 0.10$, determine whether there is enough evidence to reject the null hypothesis, and make a statement regarding the original claim.

9. **MULTIPLE CHOICE** Identify the graph that could have a correlation coefficient of -0.96 in a linear regression.

A


C


B


D


10. **DRIVING** The table lists the average number of accidents per month for sections of highway with the given speed limits.

   a. Make a scatter plot of the data, and identify the relationship.

   b. Calculate and interpret the correlation coefficient.

   c. Determine if the correlation coefficient is significant at the 5% level. Explain your reasoning.

Speed (mph)	Accidents
25	2.6
30	3.5
35	6.9
40	10.3
45	15.2
50	18.3
55	22.3
60	24.8
65	26.0
70	29.2

CHAPTER 11

Connect to AP Calculus
Population Proportions

Objective

- Create confidence intervals for population proportions.

In Lesson 11-2, you learned that the probability of a success in a single trial of a binomial experiment is p and it can be expressed as a fraction, decimal, or percentage. For example, the probability of tossing a fair coin and recording a tail is $\frac{1}{2}$, 0.5, or 50%. This probability is a *population proportion* because for the fair coin, the entire population, both heads and tails, is accounted for.

It is not always plausible to calculate population proportions. For example, calculating the percentage of high school students that own their own car would require surveying every high school student. Thus, population proportions can be estimated using *sample proportions* in the same way that sample means were used to estimate population means in this chapter.

The sample proportion $\hat{p}$ is the proportion of successes in a sample and is given by $\hat{p} = \frac{x}{n}$, where x is the number of successes in the sample and n is the sample size. The probability of failure is then given by $\hat{q} = 1 - \hat{p}$.

Activity 1 Sample Proportion

A sample of 2582 high school students found that 362 students own their own car. Estimate the population proportion of high school students that own their own car by calculating the sample proportion $\hat{p}$.

Step 1 Substitute $x = 362$ and $n = 2582$ into the formula for $\hat{p}$ and simplify.

Step 2 Interpret the result.
The percent of all high school students that own their own car is approximately 14%.

Analyze the Results

1. Is the sample proportion an accurate estimate for the population proportion? Explain your reasoning.
2. If the sample is conducted with a larger n, what can be said about the relationship between the sample proportion and the population proportion?
3. Will the sample proportion ever equal the population proportion? If not, what can be done to the sample proportion in addition to increasing n to give a better estimate for the population proportion? Explain your reasoning.

We know from Lesson 11-5 that the $\hat{p}$ found in Activity 1 is a *point estimate*. If we wanted to create a better estimate, we would want to construct an interval. The behavior of the distribution of sample proportions is similar to the distribution of sample means. As the sample size increases, the distribution becomes approximately normal and the average of the sample proportions approaches the population proportion p.

StudyTip
Normal Distribution and z-Values Recall from Lesson 11-4, that the normal distribution is used for binomial distribution when $np \geq 5$ and $nq \geq 5$. Thus we can find and use z-values to calculate E in the same way as we did in Lesson 11-5.

Just as a confidence interval can be calculated for a population mean by adding and subtracting a maximum error of estimate E to and from a sample mean $\bar{x}$ a maximum error of estimate can be added to and subtracted from a sample proportion $\hat{p}$ to create a confidence interval for a population proportion.

KeyConcept Confidence Interval For a Population Proportion

The confidence interval CI for a population proportion is given by

$$CI = \hat{p} \pm E,$$

where $\hat{p}$ is the sample proportion and E is the maximum error of estimate represented by $z\sqrt{\frac{\hat{p}\hat{q}}{n}}$.

Activity 2 Confidence Interval for a Proportion

A random survey of 825 college applicants recorded the students' high school grade point average a. Find the 90% confidence interval for the proportion of all college applicants with a grade point average of 3.0 or higher.

GPA a	Applicants
$4.0 \le a$	33
$3.0 \le a < 4.0$	600
$2.0 \le a < 3.0$	175
$a < 2.0$	17

Step 1 **Find $\hat{p}$ and $\hat{q}$.**

$\hat{p} = \frac{x}{n}$ — Sample Proportion Formula

$= \frac{633}{825}$ or about 0.77 — $x = 633$ and $n = 825$

Therefore, $\hat{q} = 1 - 0.77$ or about 0.23.

Step 2 **Verify that $n\hat{p} \ge 5$ and $n\hat{q} \ge 5$.**

$n\hat{p} \approx 825(0.77)$ or 635.25 $\quad n\hat{q} \approx (825)(0.23)$ or 189.75

Since $n\hat{p} \ge 5$ and $n\hat{q} \ge 5$, the sampling distribution of $\hat{p}$ can be approximated by the normal distribution.

Step 3 **Find the z-value.**

For a 90% confidence level, $z = 1.645$.

Step 4 **Find the maximum error of estimate.**

$E = z\sqrt{\frac{\hat{p}\hat{q}}{n}}$ — Maximum Error of Estimate Formula

$\approx 1.645\sqrt{\frac{0.77(0.23)}{825}}$ or about 0.0241 — $z = 1.645$, $\hat{p} \approx 0.77$, $\hat{q} \approx 0.23$, and $n = 825$

Step 5 **Find the left and right endpoints of the confidence interval.**

$CI = \hat{p} \pm E$ — Confidence Interval for a Proportion

$= 0.77 \pm 0.0241$ — $\hat{p} = 0.77$ and $E = 0.0241$

Left Boundary	Right Boundary
$0.77 - 0.0241 = 0.7459$	$0.77 + 0.0241 = 0.7941$

The 90% confidence interval is then $0.746 < p < 0.794$. Therefore, we are 90% confident that the proportion of applicants with a G.P.A. of 3.0 or higher is between 74.6% and 79.4%.

StudyTip

Finding z-Values Recall from Lesson 11-5 that the most common confidence levels and their corresponding z-values are as follows.

Confidence Level	z-value
90%	1.645
95%	1.960
99%	2.576

Remember that you can find the z-value for any confidence interval with a graphing calculator.

Analyze the Results

4. Describe two ways that the confidence interval found in Step 5 can be narrowed.

5. If the confidence level is held constant, what would n need to be to reduce the maximum error of estimate by $\frac{1}{2}$?

Model and Apply

6. In a 2006 Gallup Poll of 1000 adults, 480 felt that the money the government spent on the space shuttle should have been spent on something else. Find the 95% confidence interval for the proportion of all adults who felt this way.

7. A random sample of 279 households found that 58% had at least one sport-utility vehicle (SUV). Find the 99% confidence interval for the proportion of all households that own a SUV.

CHAPTER 12

Limits and Derivatives

Then

In **Chapter 1,** you learned about limits and rates of change.

Now

In Chapter 12, you will:

- Evaluate limits of polynomial and rational functions.
- Find instantaneous rates of change.
- Find and evaluate derivatives of polynomial functions.
- Approximate the area under a curve.
- Find antiderivatives, and use the Fundamental Theorem of Calculus.

Why?

BUNGEE JUMPING The basic tools of calculus, derivatives and integrals, are very useful when working with rates that are not constant. A bungee jumper experiences varying rates of descent and ascent, as well as changing acceleration, depending on her position during the jump.

PREREAD Use the Mid-Chapter Quiz to write two or three questions about the first three lessons that will help you to predict the organization of the first half of Chapter 12.

Darryl Leniuk/Riser/Getty Images