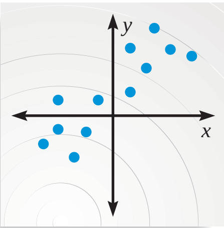
EXPLORING DATA AND STATISTICS

**2.5**

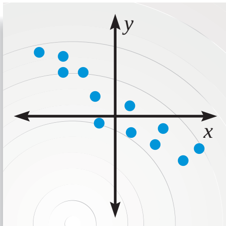
# Correlation and Best-Fitting Lines

## GOAL 1 SCATTER PLOTS AND CORRELATION

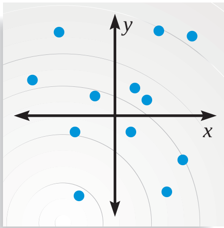
A **scatter plot** is a graph used to determine whether there is a relationship between paired data. In many real-life situations, scatter plots follow patterns that are approximately linear. If $y$ tends to increase as $x$ increases, then the paired data are said to have a **positive correlation**. If $y$ tends to decrease as $x$ increases, then the paired data are said to have a **negative correlation**. If the points show no linear pattern, then the paired data are said to have **relatively no correlation**.



**Positive correlation**



**Negative correlation**



**Relatively no correlation**

### *What* you should learn

**GOAL 1** Use a scatter plot to identify the correlation shown by a set of data.

**GOAL 2** Approximate the best-fitting line for a set of data, as applied in **Example 3**.

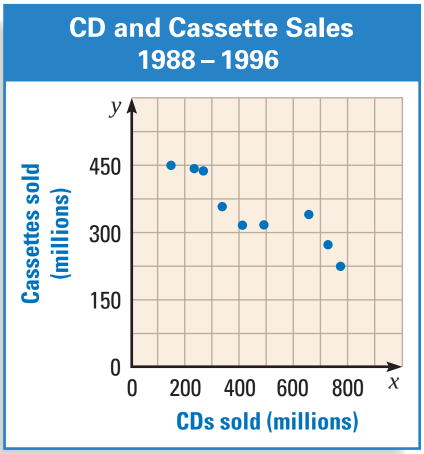
### *Why* you should learn it

▼ To identify **real-life** trends in data, such as when and for how long Old Faithful will erupt in **Ex. 23**.

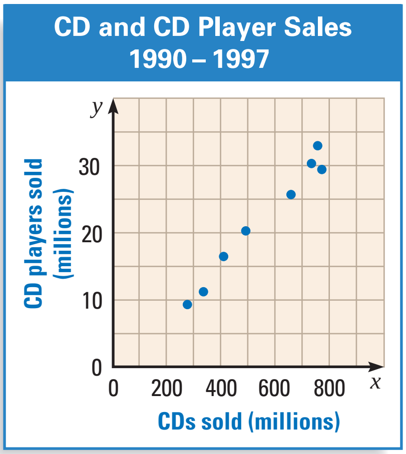
REAL LIFE

### EXAMPLE 1 *Determining Correlation*

**MUSIC** Describe the correlation shown by each scatter plot.




▶ Source: *Statistical Abstract of the United States*




▶ Sources: *Electronic Market Data Book*, Recording Industry Association of America

**SOLUTION**

The first scatter plot shows a negative correlation, which means that as CD sales increased, the sales of cassettes tended to decrease.

The second scatter plot shows a positive correlation, which means that as CD sales increased, the sales of CD players tended to increase.

## GOAL 2 APPROXIMATING BEST-FITTING LINES

When data show a positive or negative correlation, you can approximate the data with a line. Finding the line that *best* fits the data is tedious to do by hand. (See page 107 for a description of how to use technology to find the best-fitting line.) You can, however, approximate the best-fitting line using the following graphical approach.

### APPROXIMATING A BEST-FITTING LINE: GRAPHICAL APPROACH

**STEP 1** Carefully *draw a scatter plot* of the data.

**STEP 2** *Sketch the line* that appears to follow most closely the pattern given by the points. There should be as many points above the line as below it.

**STEP 3** *Choose two points* on the line, and estimate the coordinates of each point. These two points do not have to be original data points.

**STEP 4** *Find an equation of the line* that passes through the two points from Step 3. This equation models the data.

REAL LIFE

*Walking Speeds*

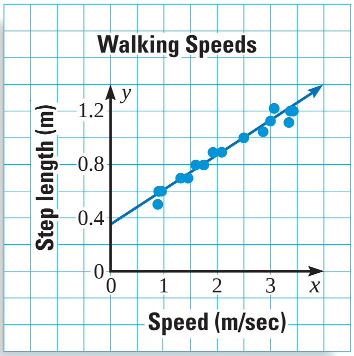
### EXAMPLE 2 *Fitting a Line to Data*

Researchers have found that as you increase your walking speed (in meters per second), you also increase the length of your step (in meters). The table gives the average walking speeds and step lengths for several people. Approximate the best-fitting line for the data.

▶ Source: *Biomechanics and Energetics of Muscular Exercise*

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Speed** | 0.8 | 0.85 | 0.9 | 1.3 | 1.4 | 1.6 | 1.75 | 1.9 |
| **Step** | 0.5 | 0.6 | 0.6 | 0.7 | 0.7 | 0.8 | 0.8 | 0.9 |
| **Speed** | 2.15 | 2.5 | 2.8 | 3.0 | 3.1 | 3.3 | 3.35 | 3.4 |
| **Step** | 0.9 | 1.0 | 1.05 | 1.15 | 1.25 | 1.15 | 1.2 | 1.2 |

**SOLUTION**




1. Begin by drawing a scatter plot of the data.
2. Next, sketch the line that appears to best fit the data.
3. Then, choose two points on the line. From the scatter plot shown, you might choose (0.9, 0.6) and (2.5, 1).
4. Finally, find an equation of the line. The line that passes through the two points has a slope of:

$$m = \frac{1 - 0.6}{2.5 - 0.9} = \frac{0.4}{1.6} = 0.25$$

Use the point-slope form to write the equation.

$y - y_1 = m(x - x_1)$ **Use point-slope form.**

$y - 0.6 = 0.25(x - 0.9)$ **Substitute for $m$, $x_1$, and $y_1$.**

$y = 0.25x + 0.375$ **Simplify.**

FOCUS ON CAREERS

**PEDIATRICIAN**
A pediatrician is a medical doctor who specializes in children's health. About 7% of all medical doctors are pediatricians.

CAREER LINK
www.mcdougallittell.com

## EXAMPLE 3 *Using a Fitted Line*

**SLEEP REQUIREMENTS** The table shows the age $t$ (in years) and the number $h$ of hours slept per day by 24 infants who were less than one year old.

| Infant Sleep Requirements | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Age, $t$** | 0.03 | 0.05 | 0.05 | 0.08 | 0.11 | 0.19 | 0.21 | 0.26 |
| **Sleep, $h$** | 15.0 | 15.8 | 16.4 | 16.2 | 14.8 | 14.7 | 14.5 | 15.4 |
| **Age, $t$** | 0.34 | 0.35 | 0.35 | 0.44 | 0.52 | 0.69 | 0.70 | 0.75 |
| **Sleep, $h$** | 15.2 | 15.3 | 14.4 | 13.9 | 14.4 | 13.2 | 14.1 | 14.2 |
| **Age, $t$** | 0.80 | 0.82 | 0.86 | 0.91 | 0.94 | 0.97 | 0.98 | 0.98 |
| **Sleep, $h$** | 13.4 | 13.2 | 13.9 | 13.1 | 13.7 | 12.7 | 13.7 | 13.6 |

**a.** Approximate the best-fitting line for the data.

**b.** Use the fitted line to estimate the number of hours that a 6 month old infant sleeps per day.

### SOLUTION

**a.** ***Draw*** a scatter plot of the data.

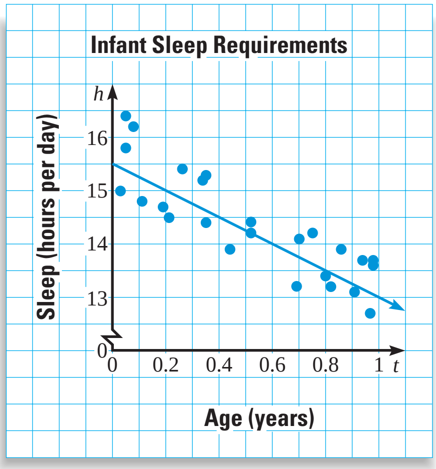
***Sketch*** the line that appears to best fit the data.

***Choose*** two points on the line. From the scatter plot shown, you might choose:

$$(0, 15.5) \text{ and } (0.52, 14.4)$$

***Find*** an equation of the line. The line that passes through the two points has a slope of:

$$m = \frac{14.4 - 15.5}{0.52 - 0} = \frac{-1.1}{0.52} \approx -2.12$$




Because the $h$-intercept was chosen as one of the two points for determining the line, you can use the slope-intercept form to approximate the best-fitting line as follows:

$h = mt + b$ — **Use slope-intercept form.**

$h = -2.12t + 15.5$ — **Substitute for $m$ and $b$.**

▶ An equation of the line is $h = -2.12t + 15.5$. Notice that a newborn infant sleeps about 15.5 hours per day and tends to sleep less as he or she gets older.

**b.** To estimate the number of hours that a 6 month old infant sleeps, use the model from part (a) and the fact that 6 months $= 0.5$ years.

$h = -2.12t + 15.5$ — **Write linear model.**

$h = -2.12(0.5) + 15.5$ — **Substitute 0.5 for $t$.**

$h \approx 14.4$ — **Simplify.**

▶ A 6 month old infant sleeps about 14.4 hours per day.

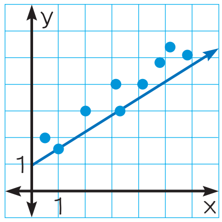
# GUIDED PRACTICE

**Vocabulary Check ✓**

**1.** Explain the meaning of the terms positive correlation, negative correlation, and relatively no correlation.

**Concept Check ✓**

**2.** Suppose you were given the shoe sizes $s$ and the heights $h$ of one hundred 25 year old men. Do you think that $s$ and $h$ would have a *positive correlation*, a *negative correlation*, or *relatively no correlation*? Explain.

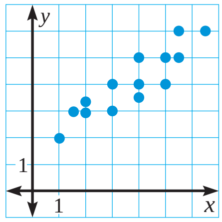


Ex. 3

**3.** **ERROR ANALYSIS** Explain why the line shown at the right is not a good fit for the data.

**Skill Check ✓**

**4.** Does the scatter plot at the right show a *positive correlation*, a *negative correlation*, or *relatively no correlation*? Explain.



Ex. 4

**5.** Look back at Example 2. Estimate the step length of a person who walks at a speed of 4 meters per second.

**FM RADIO STATIONS** **In Exercises 6 and 7, use the table below which gives the number of FM radio stations from 1989 to 1995.** ▶ Source: *Statistical Abstract of the United States*

| Years since 1989 | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| FM radio stations | 4269 | 4392 | 4570 | 4785 | 4971 | 5109 | 5730 |

**6.** Approximate the best-fitting line for the data.

**7.** If the pattern continues, how many FM radio stations will there be in 2010?

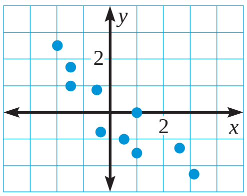
# PRACTICE AND APPLICATIONS

**STUDENT HELP**

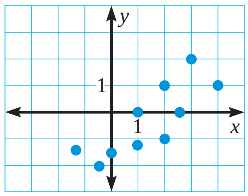
**Extra Practice** to help you master skills is on p. 942.

**DETERMINING CORRELATION** **Tell whether $x$ and $y$ have a *positive correlation*, a *negative correlation*, or *relatively no correlation*.**

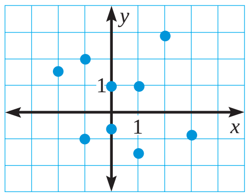
**8.**




**9.**




**10.**




**DRAWING SCATTER PLOTS** **Draw a scatter plot of the data. Then tell whether the data have a *positive correlation*, a *negative correlation*, or *relatively no correlation*.**

**STUDENT HELP**

**HOMEWORK HELP**
**Example 1:** Exs. 8–14, 22, 23
**Example 2:** Exs. 16–21
**Example 3:** Exs. 24–27

**11.**

| x | 1 | 2 | 3 | 3 | 5 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|---|
| y | 1 | 3 | 3 | 4 | 4 | 5 | 7 | 6 | 8 | 7 |

**12.**

| x | 1 | 1 | 3 | 4 | 4 | 5 | 7 | 7 | 8 | 8 |
|---|---|---|---|---|---|---|---|---|---|---|
| y | 8 | 2 | 5 | 8 | 3 | 5 | 3 | 5 | 1 | 8 |

**DRAWING SCATTER PLOTS** **Draw a scatter plot of the data. Then tell whether the data have a *positive correlation*, a *negative correlation*, or *relatively no correlation*.**

**13.**

| $x$ | 1.5 | 2 | 3 | 3.5 | 4.5 | 5 | 6 | 6.5 | 8 | 8 |
|---|---|---|---|---|---|---|---|---|---|---|
| $y$ | 7 | 8 | 6 | 7.5 | 5 | 6.5 | 3.5 | 5 | 5 | 4 |

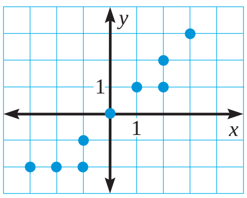
**14.**

| $x$ | 2 | 3 | 3.5 | 4 | 4.5 | 5.5 | 5.5 | 7 | 8 | 8.5 |
|---|---|---|---|---|---|---|---|---|---|---|
| $y$ | 9 | 7.5 | 7.5 | 5.5 | 6.5 | 5 | 4 | 3.5 | 2 | 1.5 |

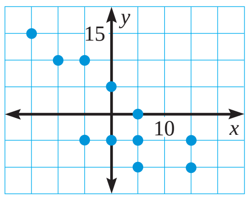
**15. LOGICAL REASONING** Explain how you can determine the type of correlation for data by examining the data in a table as opposed to drawing a scatter plot.

**APPROXIMATING BEST-FITTING LINES** **Copy the scatter plot. Then approximate the best-fitting line for the data.**

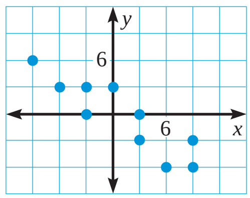
**16.**




**17.**




**18.**




**FITTING A LINE TO DATA** **Draw a scatter plot of the data. Then approximate the best-fitting line for the data.**

**19.**

| $x$ | −2 | −1 | 0 | 0.5 | 1 | 2 | 2.5 | 3.5 | 4 | 5 |
|---|---|---|---|---|---|---|---|---|---|---|
| $y$ | 4 | 3 | 2.5 | 2.5 | 2 | 0.5 | −1 | −3 | −2.5 | −2.5 |

**20.**

| $x$ | −4 | −3 | −2 | −1.5 | 0 | 0.5 | 2 | 2.5 | 3 | 4 |
|---|---|---|---|---|---|---|---|---|---|---|
| $y$ | −2 | −1 | −1.5 | 0 | 0.5 | 0.5 | 2.5 | 2 | 3 | 3 |

**21.**

| $x$ | −4 | −3 | −2 | −1.5 | −0.5 | 0 | 1 | 2 | 2.5 | 3 |
|---|---|---|---|---|---|---|---|---|---|---|
| $y$ | 6 | 4 | 4.5 | 3 | 2 | 3 | 1.5 | 2 | 0.5 | 0 |

**22. HIGH ALTITUDE TEMPERATURES** The table shows the temperature for various elevations based on a temperature of 59°F at sea level. Draw a scatter plot of the data and describe the correlation shown.

| Elevation (ft) | 1000 | 5000 | 10,000 | 15,000 | 20,000 | 30,000 |
|---|---|---|---|---|---|---|
| Temperature (°F) | 56 | 41 | 23 | 5 | −15 | −47 |

**23. OLD FAITHFUL** Old Faithful is a geyser in Yellowstone National Park. The table shows the duration of eruptions and the time interval between eruptions for a typical day. Draw a scatter plot of the data and describe the correlation shown.

| Duration (min) | 4.4 | 3.9 | 4 | 4 | 3.5 | 4.1 | 2.3 | 4.7 | 1.7 | 4.9 | 1.7 | 4.6 | 3.4 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Interval (min) | 78 | 74 | 68 | 76 | 80 | 84 | 50 | 93 | 55 | 76 | 58 | 74 | 75 |

**FOCUS ON APPLICATIONS**

REAL LIFE **OLD FAITHFUL,** shown here between eruptions, is just one of the 200–250 geysers located in Yellowstone National Park. There are more geysers and hot springs in Yellowstone than in the rest of the world combined.

STUDENT HELP

INTERNET HOMEWORK HELP
Visit our Web site www.mcdougallittell.com for help with problem solving in Exs. 24–27.

**CITY YEAR In Exercises 24 and 25, use the table below which gives the enrollment for the City Year national youth service program from 1989 to 1998.**

| Years since 1989 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|---|
| Enrollment | 57 | 76 | 107 | 234 | 371 | 688 | 678 | 716 | 894 | 918 |

**24.** Approximate the best-fitting line for the data.

**25.** If the pattern continues, how many people will enroll in City Year in 2010?

**BIOLOGY CONNECTION In Exercises 26 and 27, use the table below which gives the average life expectancy (in years) of a person based on various years of birth.** ▶ Source: National Center for Health Statistics

| Year of birth | 1900 | 1910 | 1920 | 1930 | 1940 |
|---|---|---|---|---|---|
| Life expectancy | 47.3 | 50 | 54.1 | 59.7 | 62.9 |
| Year of birth | 1950 | 1960 | 1970 | 1980 | 1990 |
| Life expectancy | 68.2 | 69.7 | 70.8 | 73.7 | 75.4 |

**26.** Approximate the best-fitting line for the data.

**27.** Predict the life expectancy for someone born in 2010.

Test Preparation

**28.** **MULTI-STEP PROBLEM** The table below gives the numbers (in thousands) of black-and-white and color televisions sold in the United States for various years from 1955–1995. ▶ Source: Electronic Industries Association

| Year | Black-and-white TVs sold (thousands) | Color TVs sold (thousands) |
|---|---|---|
| 1955 | 7,738 | 20 |
| 1960 | 5,709 | 120 |
| 1965 | 8,753 | 2,694 |
| 1970 | 4,704 | 5,320 |
| 1975 | 4,955 | 6,486 |
| 1980 | 6,684 | 10,897 |
| 1985 | 3,684 | 16,995 |
| 1990 | 1,411 | 20,384 |
| 1995 | 480 | 25,600 |

**a.** Draw a scatter plot of the data pairs (*year*, *black-and-white TVs sold*). Then describe the correlation shown by the scatter plot.

**b.** Draw a scatter plot of the data pairs (*year*, *color TVs sold*). Then describe the correlation shown by the scatter plot.

**c.** **CRITICAL THINKING** Based on your answers to parts (a) and (b), are black-and-white television sales and color television sales *positively correlated*, *negatively correlated*, or *neither*? Explain.

★ Challenge

**29.** **BEST-FITTING LINES** Describe a set of real-life data where the best-fitting line could *not* be used to make a prediction. Explain.

# MIXED REVIEW

**SOLVING INEQUALITIES** **Solve the inequality. Then graph your solution.** **(Review 1.6 for 2.6)**

**30.** $2x - 9 \geq 14$

**31.** $3(x + 7) < -x + 10$

**32.** $17 \leq 2x - 7 \leq 29$

**33.** $x - 4 < 0$ or $x - 6 \geq 4$

**DETERMINING STEEPNESS** **Tell which line is steeper.** **(Review 2.2)**

**34.** Line 1: through $(-3, 4)$ and $(1, 6)$
Line 2: through $(1, -5)$ and $(6, 2)$

**35.** Line 1: through $(6, 1)$ and $(-4, 4)$
Line 2: through $(-2, 3)$ and $(1, -6)$

**36.** Line 1: through $(2, 4)$ and $(1, 7)$
Line 2: through $(-5, 8)$ and $(3, 8)$

**37.** Line 1: through $(4, 3)$ and $(1, -9)$
Line 2: through $(-2, -4)$ and $(3, -7)$

**GRAPHING EQUATIONS** **Graph the equation.** **(Review 2.3 for 2.6)**

**38.** $y = \frac{1}{3}x + 5$

**39.** $y = -10x + 9$

**40.** $y = \frac{7}{3}$

**41.** $-x + 2y = -8$

**42.** $4x + 2y = 1$

**43.** $x = 12$

# QUIZ 2

***Self-Test for Lessons 2.4 and 2.5***

**Write an equation of the line that passes through the given point and has the given slope.** **(Lesson 2.4)**

**1.** $(0, 6)$, $m = \frac{2}{3}$

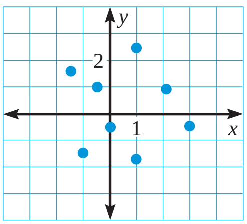
**2.** $(-4, -3)$, $m = 2$

**3.** $(2, -7)$, $m = -\frac{1}{5}$

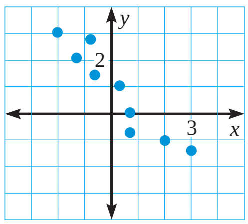
**4.** Write an equation of the line that passes through $(1, -2)$ and is perpendicular to the line that passes through $(4, 2)$ and $(0, 4)$. **(Lesson 2.4)**

**Tell whether *x* and *y* have a *positive correlation*, a *negative correlation*, or *relatively no correlation*.** **(Lesson 2.5)**

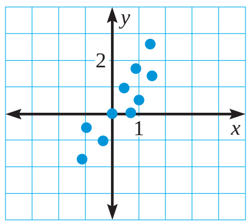
**5.**




**6.**




**7.**




**8.** **WAVES** The water depth $d$ (in feet) at which a wave breaks varies directly with the height $h$ (in feet) of the wave. A 6.5 foot wave breaks at a water depth of 8.45 feet. Write a linear model that gives $d$ as a function of $h$. If a wave breaks at a depth of 5.2 feet, what is its height? **(Lesson 2.4)**

**9.** **HEIGHTS OF CHILDREN** The table gives the average heights of children for ages 1–10. Draw a scatter plot of the data and approximate the best-fitting line for the data. **(Lesson 2.5)**

| Age (years) | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Height (cm) | 73 | 85 | 93 | 100 | 107 | 113 | 120 | 124 | 130 | 135 |