EXPLORING DATA AND STATISTICS

5.4

Fitting a Line to Data

GOAL 1 FITTING A LINE TO DATA

What you should learn

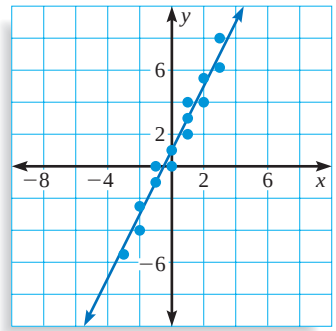
GOAL 1 Find a linear equation that approximates a set of data points.

GOAL 2 Determine whether there is a positive or negative correlation or no correlation in a set of **real-life** data, like the Olympic data in **Example 3.**

Why you should learn it

To investigate trends in **real-life** data and to make predictions such as future football salaries in **Exs. 23** and **24.**

In this lesson you will learn how to write a linear model to represent a collection of data points.



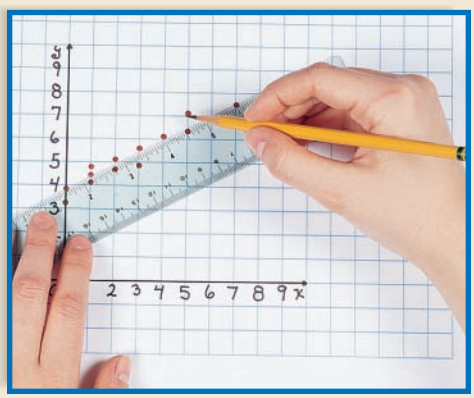
Usually there is no single line that passes through all of the data points, so you try to find the line that best fits the data, as shown at the right. This is called the **best-fitting line**.

There is a mathematical definition of the best-fitting line that is called *least squares approximation*. Many calculators have a built-in program for finding the equation of the best-fitting line. Since least squares approximation is not part of Algebra 1, you will be asked to use a graphical approach for drawing a line that is probably close to the best-fitting line.

ACTIVITY
Developing Concepts

Approximating a Best-Fitting Line

With your group, use the following steps to approximate a best-fitting line.



1. Carefully plot the following points on graph paper.

   (0, 3.3), (0, 3.9), (1, 4.2), (1, 4.5), (1, 4.8), (2, 4.7), (2, 5.1), (3, 4.9), (3, 5.6), (4, 6.1), (5, 6.4), (5, 7.1), (6, 6.8), (7, 7.5), (8, 7.8)

2. Use a ruler to sketch the line that you think best approximates the data points. Describe your strategy.

3. Locate two points on the line. Approximate the x-coordinate and the y-coordinate for each point. (These do not have to be two of the original data points.)

4. Use the method from Lesson 5.3 to find an equation of the line that passes through the two points.

REAL LIFE Biology

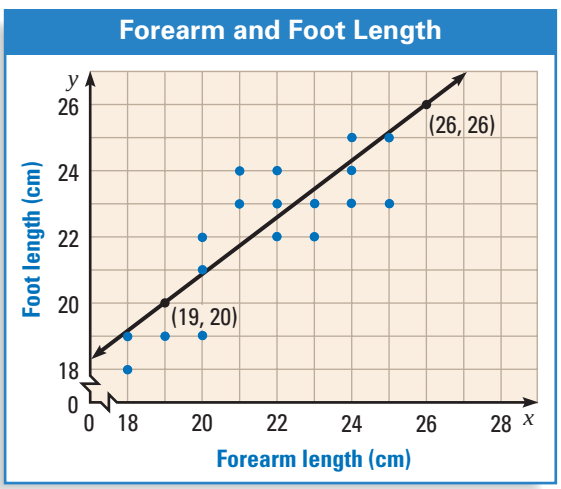
EXAMPLE 1 Approximating a Best-Fitting Line

The data in the table show the forearm lengths and foot lengths (without shoes) of 18 students in an algebra class. After graphing these data points, draw a line that corresponds closely to the data. Write an equation of your line.

Forearm length	Foot length
22 cm	24 cm
20 cm	19 cm
24 cm	24 cm
21 cm	23 cm
25 cm	23 cm
18 cm	18 cm
20 cm	21 cm
23 cm	23 cm
24 cm	25 cm
20 cm	22 cm
19 cm	19 cm
25 cm	25 cm
23 cm	22 cm
22 cm	23 cm
18 cm	19 cm
24 cm	23 cm
21 cm	24 cm
22 cm	22 cm

SOLUTION

Let x represent the forearm length and let y represent the foot length. To begin, plot the points given by the ordered pairs. Then sketch the line that appears to best fit the points.



Next, find two points that lie on the line. You might choose the points (19, 20) and (26, 26). Find the slope of the line through these two points.

$$m = \frac{y_2 - y_1}{x_2 - x_1}$$ Write slope formula.

$$m = \frac{26 - 20}{26 - 19}$$ Substitute.

$$m = \frac{6}{7}$$ Simplify.

$$\approx 0.86$$ Decimal approximation.

To find the y-intercept of the line, substitute the values $m = 0.86$, $x = 19$, and $y = 20$ in the slope-intercept form.

$$y = mx + b$$ Write slope-intercept form.

$$20 = (0.86)(19) + b$$ Substitute 0.86 for m, 19 for x, and 20 for y.

$$20 = 16.34 + b$$ Simplify.

$$3.66 = b$$ Solve for b.

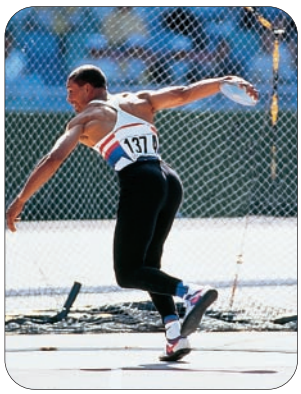
An approximate equation of the best-fitting line is $y = 0.86x + 3.66$. In general, if a student has a long forearm, then that student also has a long foot.

STUDENT HELP

Study Tip

If you choose different points, you might get a different linear model. Be sure to choose points that will give you a line that is close to most of the points.

FOCUS ON APPLICATIONS



REAL LIFE DISCUS EVENT The discus throw is one of the original Olympic events. New records for the distance an athlete can throw the discus continue to be set.

EXAMPLE 2 *Approximating a Best-Fitting Line*

DISCUS THROWS The winning Olympic discus throws from 1908 to 1996 are shown in the table. After graphing these data points, draw a line that corresponds closely to the data. Write an equation of your line.

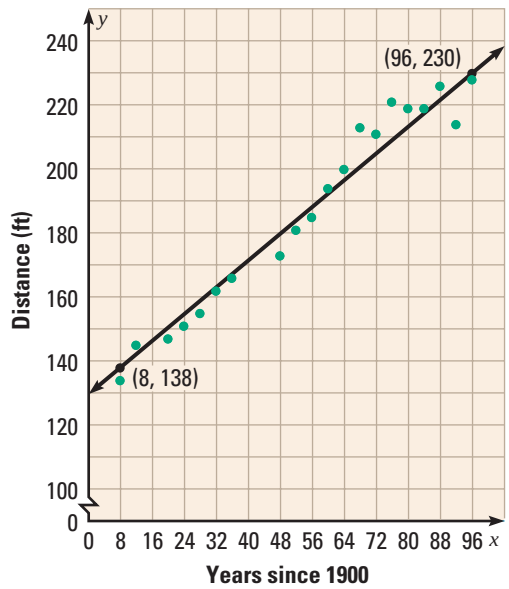
Olympic year	Winning throw
1908	134.1 ft
1912	148.3 ft
1920	146.6 ft
1924	151.4 ft
1928	155.3 ft
1932	162.4 ft
1936	165.6 ft
1948	173.2 ft
1952	180.5 ft
1956	184.9 ft
1960	194.2 ft
1964	200.1 ft
1968	212.5 ft
1972	211.3 ft
1976	221.5 ft
1980	218.7 ft
1984	218.5 ft
1988	225.8 ft
1992	213.7 ft
1996	227.7 ft

INTERNET **DATA UPDATE** of *Information Please Almanac* at www.mcdougallittell.com

SOLUTION

Let x represent the years since 1900. Let y represent the winning throw. To begin, plot the points given by the ordered pairs. Then sketch the line that appears to best fit the points.

Winning Discus Throws



Next, find two points that lie on the line, such as (8, 138) and (96, 230). Find the slope of the line through these points.

$$m = \frac{y_2 - y_1}{x_2 - x_1} = \frac{230 - 138}{96 - 8} = \frac{92}{88} \approx 1.05$$

To find the y-intercept of the line, substitute the values $m = 1.05$, $x = 8$, and $y = 138$ in the slope-intercept form.

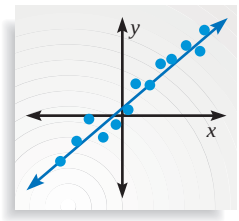
$y = mx + b$	Write slope-intercept form.
$138 = (1.05)(8) + b$	Substitute 1.05 for m, 8 for x, and 138 for y.
$138 = 8.4 + b$	Simplify.
$129.6 = b$	Solve for b.

▶ An approximate equation of the best-fitting line is $y = 1.05x + 129.6$. In most years, the winner of the discus throw was able to throw the discus farther than the previous winner.

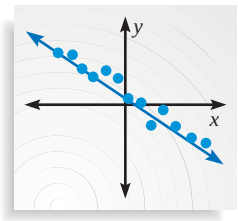
GOAL 2 DETERMINING THE CORRELATION OF x AND y

Correlation is a number r satisfying $-1 \leq r \leq 1$ that indicates how well a particular set of data can be approximated by a straight line. The Activity on page 299 shows how to find an r-value using a calculator. In this course you will describe a correlation without actually finding r-values, as described below.

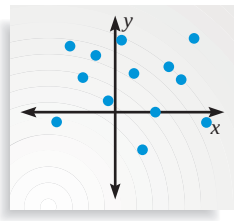
When the points on a scatter plot can be approximated by a line with positive slope, x and y have a **positive correlation**. When the points can be approximated by a line with negative slope, x and y have a **negative correlation**. When the points cannot be well approximated by a straight line, we say that there is **relatively no correlation.**



Positive correlation



Negative correlation



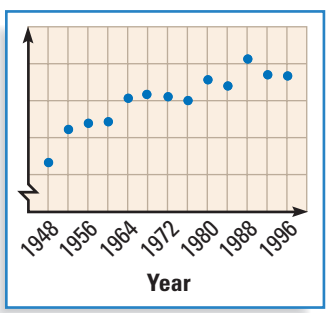
Relatively no correlation

EXAMPLE 3 *Using Correlation*

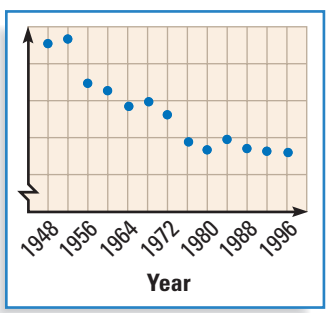
In swimming events, performance is positively correlated with time.

a. The two graphs show the winning 100-meter women's freestyle swimming times and the winning women's long jump distances for the Olympics from 1948 through 1996. Which is which? Explain your reasoning.

b. Describe the correlation of each set of data.



▶ Source: *Information Please Almanac*



▶ Source: *Information Please Almanac*

FOCUS ON APPLICATIONS

REAL LIFE **FREESTYLE SWIMMING** In a freestyle swimming event, the swimmer can choose the stroke. Usually, the swimmer chooses the Australian Crawl.

INTERNET **APPLICATION LINK** www.mcdougallittell.com

SOLUTION

a. The first graph must represent the long jump distances because the winners have tended to jump farther with each Olympic year. The second graph must represent the swimming times because the winners have tended to swim faster with each Olympic year, so their times have been decreasing.

b. The first graph shows a positive correlation between the year and the winning distance. The second graph shows a negative correlation between the year and the winning time.

GUIDED PRACTICE

Vocabulary Check ✓

1. The line that most closely fits a set of data is called the ? .

Concept Check ✓

Draw a scatter plot of data that have the given correlation.

2. Positive **3.** Negative **4.** Relatively no

5. **DISCUS THROWS** Predict the distance of the winning throw in the year 2005 from the data in Example 2. Explain your reasoning.

Skill Check ✓

Draw a scatter plot of the data. State whether *x* and *y* have a *positive correlation*, a *negative correlation*, or *relatively no correlation*. If possible, draw a line that closely fits the data and write an equation of the line.

6.

x	y
1	2
2	9
3	8
4	1
5	4
6	8

7.

x	y
−3	8
−2	6
−1	5
0	3
1	2
2	0

8.

x	y
1.1	5.1
1.7	5.5
2.2	5.9
2.6	6.3
3.3	7.5
3.5	7.6

9.

x	y
5.5	0.4
6.2	1.0
7.7	2.5
8.1	2.9
9.2	4.3
9.7	5.5

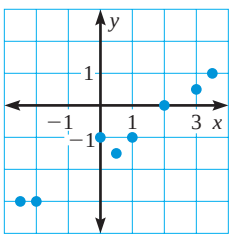
PRACTICE AND APPLICATIONS

STUDENT HELP

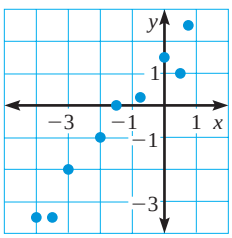
Extra Practice to help you master skills is on p. 801.

FITTING LINES **Copy the graph and draw a line that corresponds closely to the data. Write an equation of your line.**

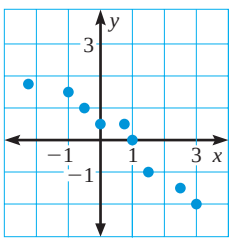
10.



11.



12.



FITTING LINES **Draw a scatter plot of the data. Draw a line that corresponds closely to the data and write an equation of the line.**

13.

x	y
1.0	3.8
1.5	4.2
1.7	5.3
2.0	5.8
2.0	5.5
1.5	6.7

14.

x	y
3.0	7.1
3.4	8.1
4.0	8.5
4.1	8.9
4.8	9.6
5.2	9.8

15.

x	y
3.0	9.9
3.5	9.7
3.7	8.6
4.0	8.1
4.0	8.4
4.5	7.4

16.

x	y
5.0	6.8
5.4	5.8
6.0	5.6
6.1	5.2
6.8	4.3
7.2	3.5

STUDENT HELP

HOMEWORK HELP

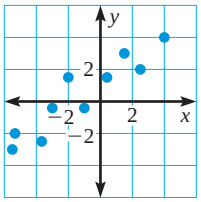
Example 1: Exs. 10–16, 23

Example 2: Exs. 10–16, 23

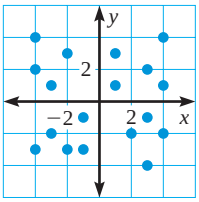
Example 3: Exs. 17–22

GRAPHICAL REASONING **State whether *x* and *y* have a *positive correlation*, a *negative correlation*, or *relatively no correlation*.**

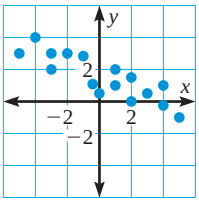
17.



18.



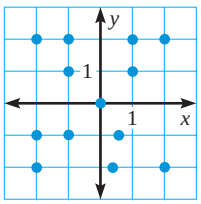
19.



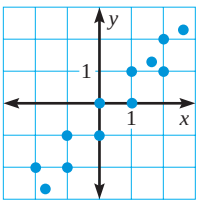
STUDENT HELP

HOMEWORK HELP
Visit our Web site www.mcdougallittell.com for help with Exs. 23 and 24.

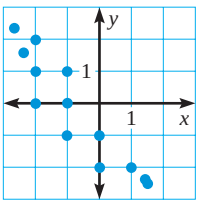
20.



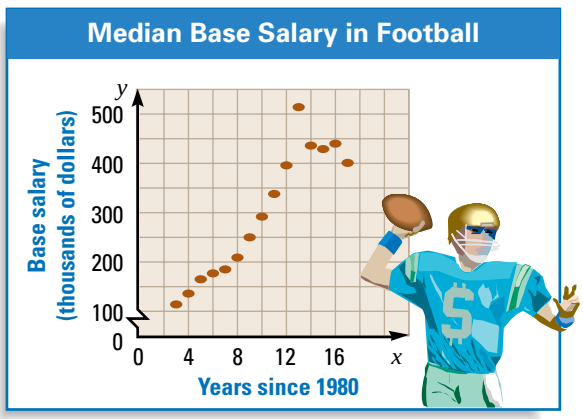
21.



22.



FOOTBALL SALARIES
In Exercises 23 and 24, use the following information.
The median base salary for players in the National Football League from 1983 to 1997 is shown in the scatter plot at the right. In the scatter plot, *y* represents the salary and *x* represents the number of years since 1980.



▶ Source: National Football League Players Association

23. Find an equation of the line that you think closely fits the data.

24. Use the equation from Exercise 23 to approximate the median base salary in the year 2010.

BIOLOGY CONNECTION **In Exercises 25–27, use the following information.**
As people grow older, the size of their pupils tends to get smaller. The average diameter (in millimeters) of one person's pupils is given in the table.

Sample Pupil Diameters		
Age (years)	Day	Night
20	4.7	8.0
30	4.3	7.0
40	3.9	6.0
50	3.5	5.0
60	3.1	4.1
70	2.7	3.2
80	2.3	2.5

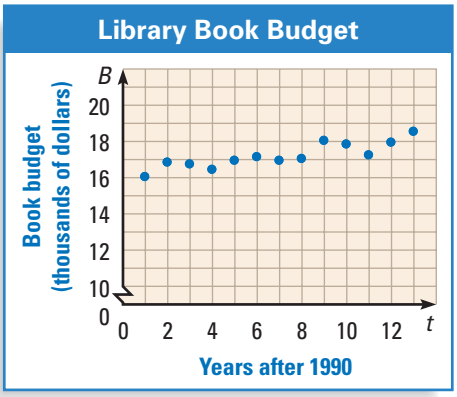
25. Draw a scatter plot of the day diameters and another of the night diameters. Let *x* represent the person's age and let *y* represent pupil diameters.

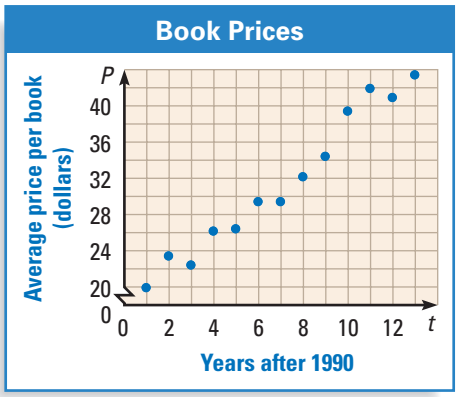
26. Find an equation of the line that closely fits the day and the night sets of data for pupil diameters.

27. **CRITICAL THINKING** Do the two lines have the same slope? Explain your answer in the context of the real-life data.

Test Preparation

MULTI-STEP PROBLEM **In Exercises 28–31, use the scatter plots. They show a library's annual book budget *B* and the average price *P* of a book.**





28. Find an equation of the line that you think best represents the library's budget for purchasing books.

29. Find an equation of the line that you think best represents the average price of a new book purchased by the library.

30. Interpret the slopes of the two lines in the context of the problem.

31. **CRITICAL THINKING** Suppose that you want to ask town officials to increase the library's annual book budget. Use your results and the information above to write a letter that explains why the library budget should be increased.

★ Challenge

ANALYZING DATA **In Exercises 32–34, use the Book Prices scatter plot above.**

32. Explain why you would not want to choose the points for the years 1992 and 1997 to write an equation that represents the data.

33. Will choosing the two points that are the farthest apart always give you the closest line? Explain why or why not. If not, sketch a counterexample.

EXTRA CHALLENGE
www.mcdougallittell.com

34. Explain how to choose a good pair of points to find a line that is probably close to the best-fitting line.

MIXED REVIEW

HORIZONTAL AND VERTICAL LINES **Decide whether the line is *horizontal* or *vertical*. Then graph the line.** (Review 4.2)

35. $y = -2$ **36.** $y = 3$ **37.** $x = 4$ **38.** $x = -5$

VISUAL REASONING **Without calculating, state whether the slope of the line through the points is *positive, negative, zero,* or *undefined.*** (Review 4.4)

39. (1, 4), (3, 5) **40.** (2, 5), (4, 1) **41.** (6, 1), (3, 1) **42.** (5, 2), (4, 3)

WRITING EQUATIONS OF LINES **Write an equation of the line that passes through the points.** (Review 5.3 for 5.5)

43. (2, 5), (6, 4) **44.** (1, 4), (3, 7) **45.** (3, 7), (7, 3) **46.** (5, 2), (4, 3)

47. (−3, 1), (4, −2) **48.** (−2, −3), (0, 4) **49.** (5, 1), (3, −6) **50.** (0, −6), (−1, 7)