4.6 Fit a Line to Data

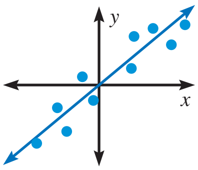
Before You modeled situations involving a constant rate of change.
Now You will make scatter plots and write equations to model data.
Why? So you can model scientific data, as in Ex. 19.

Key Vocabulary
- scatter plot
- correlation
- line of fit

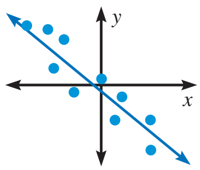
COMMON CORE

CC.9-12.S.ID.6c Fit a linear function for a scatter plot that suggests a linear association.*

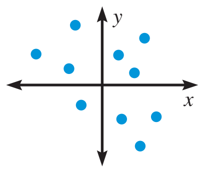
A **scatter plot** is a graph used to determine whether there is a relationship between paired data. Scatter plots can show trends in the data.



If y tends to increase as x increases, the paired data are said to have a **positive correlation**.



If y tends to decrease as x increases, the paired data are said to have a **negative correlation**.

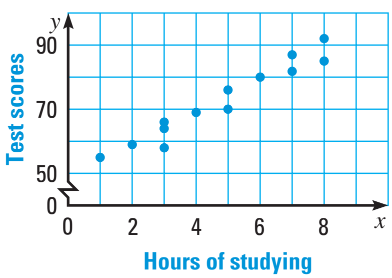


If x and y have no apparent relationship, the paired data are said to have **relatively no correlation**.

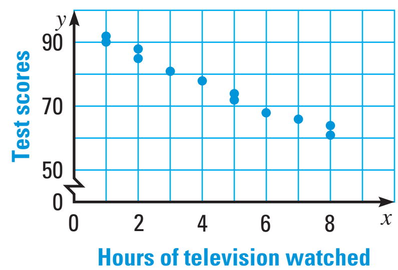
EXAMPLE 1 Describe the correlation of data

Describe **the correlation of the data graphed in the scatter plot.**

a.



b.



a. The scatter plot shows a positive correlation between hours of studying and test scores. This means that as the hours of studying increased, the test scores tended to increase.

b. The scatter plot shows a negative correlation between hours of television watched and test scores. This means that as the hours of television watched increased, the test scores tended to decrease.

✓ GUIDED PRACTICE for Example 1

1. Using the scatter plots in Example 1, predict a reasonable test score for 4.5 hours of studying and 4.5 hours of television watched.

EXAMPLE 2 Make a scatter plot

SWIMMING SPEEDS The table shows the lengths (in centimeters) and swimming speeds (in centimeters per second) of six fish.

Fish	Pike	Red gurnard	Black bass	Gurnard	Norway haddock
Length (cm)	37.8	19.2	21.3	26.2	26.8
Speed (cm/sec)	148	47	88	131	98

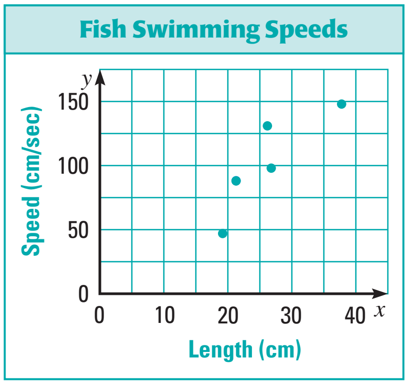
a. Make a scatter plot of the data.

b. *Describe* the correlation of the data.

Solution

a. Treat the data as ordered pairs. Let x represent the fish length (in centimeters), and let y represent the speed (in centimeters per second). Plot the ordered pairs as points in a coordinate plane.

b. The scatter plot shows a positive correlation, which means that longer fish tend to swim faster.



✓ GUIDED PRACTICE for Example 2

2. Make a scatter plot of the data in the table. *Describe* the correlation of the data.

x	1	1	2	3	3	4	5	5	6
y	2	3	4	4	5	5	5	7	8

MODELING DATA When data show a positive or negative correlation, you can model the trend in the data using a **line of fit**.

KEY CONCEPT — *For Your Notebook*

Using a Line of Fit to Model Data

STEP 1 **Make** a scatter plot of the data.

STEP 2 **Decide** whether the data can be modeled by a line.

STEP 3 **Draw** a line that appears to fit the data closely. There should be approximately as many points above the line as below it.

STEP 4 **Write** an equation using two points on the line. The points do not have to represent actual data pairs, but they must lie on the line of fit.

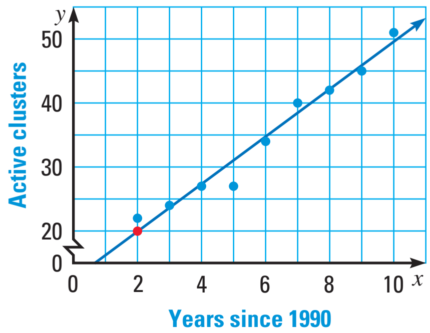
EXAMPLE 3 Write an equation to model data

BIRD POPULATIONS The table shows the number of active red-cockaded woodpecker clusters in a part of the De Soto National Forest in Mississippi. Write an equation that models the number of active clusters as a function of the number of years since 1990.

Year	1992	1993	1994	1995	1996	1997	1998	1999	2000
Active clusters	22	24	27	27	34	40	42	45	51

Solution

STEP 1 **Make** a scatter plot of the data. Let x represent the number of years since 1990. Let y represent the number of active clusters.



STEP 2 **Decide** whether the data can be modeled by a line. Because the scatter plot shows a positive correlation, you can fit a line to the data.

STEP 3 **Draw** a line that appears to fit the points in the scatter plot closely.

STEP 4 **Write** an equation using two points on the line. Use (2, 20) and (8, 42).

Find the slope of the line.

$$m = \frac{y_2 - y_1}{x_2 - x_1} = \frac{42 - 20}{8 - 2} = \frac{22}{6} = \frac{11}{3}$$

Find the y-intercept of the line. Use the point (2, 20).

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

$20 = \frac{11}{3}(2) + b$ — **Substitute $\frac{11}{3}$ for m, 2 for x, and 20 for y.**

$\frac{38}{3} = b$ — **Solve for b.**

An equation of the line of fit is $y = \frac{11}{3}x + \frac{38}{3}$.

▶ The number y of active woodpecker clusters can be modeled by the function $y = \frac{11}{3}x + \frac{38}{3}$ where x is the number of years since 1990.

Animated Algebra at my.hrw.com

✓ GUIDED PRACTICE for Example 3

3. Use the data in the table to write an equation that models y as a function of x.

x	1	2	3	4	5	6	8
y	3	5	8	9	11	12	14

EXAMPLE 4 Interpret a model

Refer to the model for the number of woodpecker clusters in Example 3.

a. *Describe* the domain and range of the function.

b. At about what rate did the number of active woodpecker clusters change during the period 1992–2000?

Solution

a. The domain of the function is the the period from 1992 to 2000, or $2 \le x \le 10$. The range is the the number of active clusters given by the function for $2 \le x \le 10$, or $20 \le y \le 49.3$.

b. The number of active woodpecker clusters increased at a rate of $\frac{11}{3}$ or about 3.7 woodpecker clusters per year.

✓ GUIDED PRACTICE for Example 4

4. In Guided Practice Exercise 2, at about what rate does y change with respect to x?

4.6 EXERCISES

HOMEWORK KEY

◯ = See **WORKED-OUT SOLUTIONS** Exs. 7 and 17

★ = **STANDARDIZED TEST PRACTICE** Exs. 2, 8, 11, 12, and 16

SKILL PRACTICE

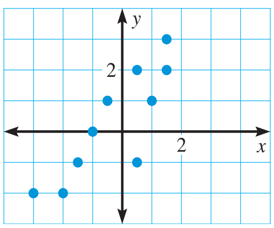
1. VOCABULARY Copy and complete: When data have a positive correlation, the dependent variable tends to __?__ as the independent variable increases.

2. ★ WRITING *Describe* how paired data with a positive correlation, a negative correlation, and relatively no correlation differ.

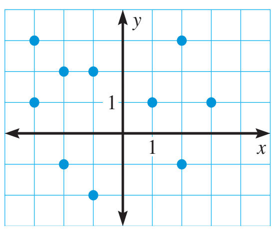
DESCRIBING CORRELATIONS **Tell whether x and y show a *positive correlation*, a *negative correlation*, or *relatively no correlation*.**

EXAMPLE 1 for Exs. 3–5, 10, 11

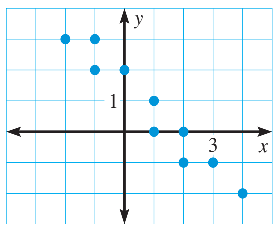
3.



4.



5.



EXAMPLES 2 and 3 for Exs. 6–9

FITTING LINES TO DATA **Make a scatter plot of the data in the table. Draw a line of fit. Write an equation of the line.**

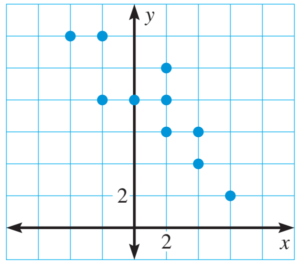
6.

x	1	1	3	4	5	6	9
y	10	12	33	46	59	70	102

7.

x	1.2	1.8	2.3	3.0	4.4	5.2
y	10	7	5	−1	−4	−8

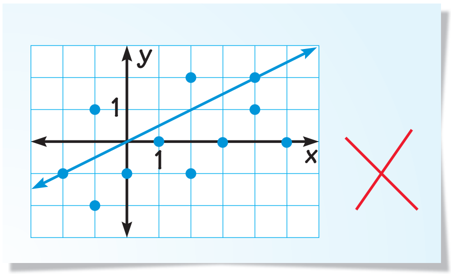
8. ★ **MULTIPLE CHOICE** Which equation best models the data in the scatter plot?



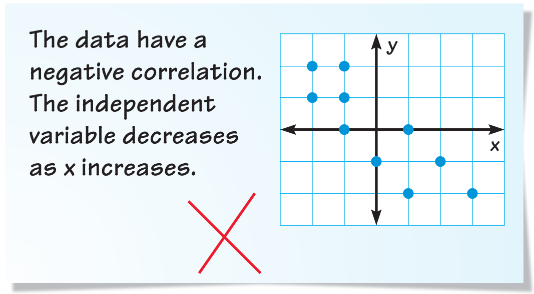
Ⓐ $y = -x - 6$ Ⓑ $y = x - 6$

Ⓒ $y = -x + 8$ Ⓓ $y = x + 8$

9. **ERROR ANALYSIS** *Describe* and correct the error in fitting the line to the data in the scatter plot.



10. **ERROR ANALYSIS** *Describe* and correct the error in describing the correlation of the data in the scatter plot.



11. ★ **OPEN-ENDED** Give an example of a data set that shows a negative correlation.

12. ★ **SHORT RESPONSE** Make a scatter plot of the data. *Describe* the correlation of the data. Is it possible to fit a line to the data? If so, write an equation of the line. If not, explain why.

x	−12	−7	−4	−3	−1	2	5	6	7	9	15
y	150	50	15	10	1	5	22	37	52	90	226

MODELING DATA Make a scatter plot of the data. *Describe* the correlation of the data. If possible, fit a line to the data and write an equation of the line.

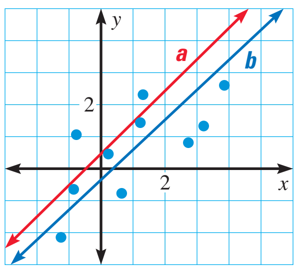
13.

x	10	12	15	20	30	45	60	99
y	−2	4	9	16	32	55	87	128

14.

x	−5	−3	−3	0	1	2	5	6
y	−4	12	10	−6	8	0	3	−9

15. **CHALLENGE** Which line shown is a better line of fit for the scatter plot? *Explain* your reasoning.



PROBLEM SOLVING

EXAMPLE 2 for Exs. 16

16. ★ **SHORT RESPONSE** The table shows the approximate home range size of big cats (members of the *Panthera* genus) in their natural habitat and the percent of time that the cats spend pacing in captivity.

Big cat (*Panthera* genus)	Lion	Jaguar	Leopard	Tiger
Home range size (km^2)	148	90	34	48
Pacing (percent of time)	48	21	11	16

a. Make a scatter plot of the data.

b. *Describe* the correlation of the data.

c. The snow leopard's home range size is about 39 square kilometers. It paces about 7% of its time in captivity. Does the snow leopard fit the pacing trend of cats in the *Panthera* genus? *Explain* your reasoning.

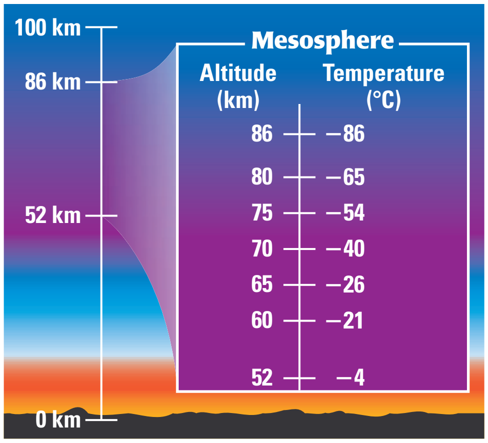
EXAMPLES 3 and 4 for Exs. 17–18

17. **EARTH SCIENCE** The mesosphere is a layer of atmosphere that lies from about 50 kilometers above Earth's surface to about 90 kilometers above Earth's surface. The diagram shows the temperature at certain altitudes in the mesosphere.

a. Make a scatter plot of the data.

b. Write an equation that models the temperature (in degrees Celsius) as a function of the altitude (in kilometers) above 50 kilometers.

c. At about what rate does the temperature change with increasing altitude in the mesosphere?



18. **ALLIGATORS** The table shows the weights of two alligators at various times during a feeding trial. Make two scatter plots, one for each alligator, where x is the number of weeks and y is the weight of the alligator. Draw lines of fit for both scatter plots. *Compare* the approximate growth rates.

Weeks	0	9	18	27	34	43	49
Alligator 1 weight (pounds)	6	8.6	10	13.6	15	17.2	19.8
Alligator 2 weight (pounds)	6	9.2	12.8	13.6	20.2	21.4	24.3

19. **GEOLOGY** The table shows the duration of several eruptions of the geyser Old Faithful and the interval between eruptions. Write an equation that models the interval as a function of an eruption's duration.

Duration (minutes)	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
Interval (minutes)	50	57	65	71	76	82	89	95

20. DAYLIGHT The table shows the number of hours and minutes of daylight in Baltimore, Maryland, for ten days in January.

Day in January	5	6	7	8	9	10	11	12	13	14
Daylight (hours and minutes)	9:30	9:31	9:32	9:34	9:35	9:36	9:37	9:38	9:40	9:41

a. Write an equation that models the hours of daylight (in minutes in excess of 9 hours) as a function of the number of days since January 5.

b. At what rate do the hours of daylight change over time in early January?

c. Do you expect the trend described by the equation to continue indefinitely? *Explain.*

21. CHALLENGE The table shows the estimated amount of time and the estimated amount of money the average person in the U.S. spent on the Internet each year from 1999 to 2005.

Year	1999	2000	2001	2002	2003	2004	2005
Internet time (hours)	88	107	136	154	169	182	193
Internet spending (dollars)	40.55	49.64	68.70	84.73	97.76	110.46	122.67

a. Write an equation that models the amount of time h (in hours) spent on the Internet as a function of the number of years y since 1999.

b. Write an equation that models the amount of money m spent on the Internet as a function of the time h (in hours) spent on the Internet.

c. Substitute the expression that is equal to h from part (a) in the function from part (b). What does the new function tell you?

d. Does the function from part (c) agree with the data given? *Explain.*

Graphing Calculator ACTIVITY *Use after Fit a Line to Data*

my.hrw.com
Keystrokes

Perform Linear Regression

MATHEMATICAL PRACTICES

Use appropriate tools strategically.

QUESTION How can you model data with the best-fitting line?

The line that most closely follows a trend in data is the *best-fitting line*. The process of finding the best-fitting line to model a set of data is called *linear regression*. This process can be tedious to perform by hand, but you can use a graphing calculator to make a scatter plot and perform linear regression on a data set.

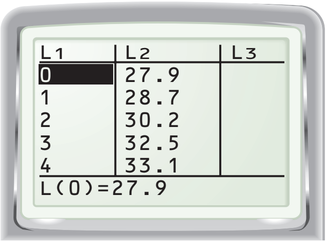
EXAMPLE 1 Create a scatter plot

The table shows the total sales from women's clothing stores in the United States from 1997 to 2002. Make a scatter plot of the data. *Describe* the correlation of the data.

Year	1997	1998	1999	2000	2001	2002
Sales (billions of dollars)	27.9	28.7	30.2	32.5	33.1	34.3

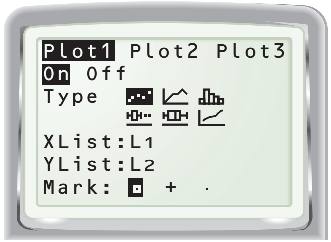
STEP 1 Enter data

Press STAT and select Edit. Enter years since 1997 (0, 1, 2, 3, 4, 5) into List 1 (L_1). These will be the x-values. Enter sales (in billions of dollars) into List 2 (L_2). These will be the y-values.



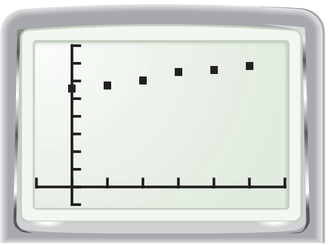
STEP 2 Choose plot settings

Press 2nd Y= and select Plot1. Turn Plot1 On. Select scatter plot as the type of display. Enter L_1 for the Xlist and L_2 for the Ylist.



STEP 3 Make a scatter plot

Press ZOOM 9 to display the scatter plot so that the points for all data pairs are visible.



STEP 4 Describe the correlation

Describe the correlation of the data in the scatter plot.

The data have a positive correlation. This means that with each passing year, the sales of women's clothing tended to increase.

MODELING DATA The *correlation coefficient* r for a set of paired data measures how well the best-fitting line fits the data. You can use a graphing calculator to find a value for r.

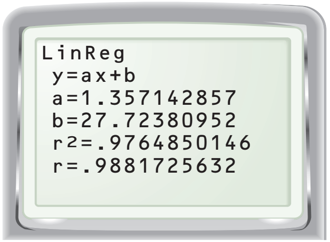
For r close to 1, the data have a strong positive correlation. For r close to -1, the data have a strong negative correlation. For r close to 0, the data have relatively no correlation.

EXAMPLE 2 Find the best-fitting line

Find an equation of the best-fitting line for the scatter plot from Example 1. Determine the correlation coefficient of the data. Graph the best-fitting line.

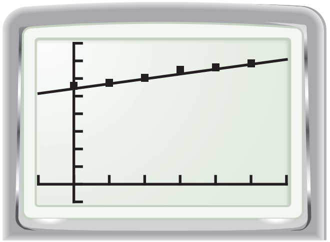
STEP 1 *Perform regression*

Press STAT. From the CALC menu, choose LinReg(ax+b). The a- and b-values given are for an equation of the form $y = ax + b$. Rounding these values gives the equation $y = 1.36x + 27.7$. Because r is close to 1, the data have a strong positive correlation.



STEP 2 *Draw the best-fitting line*

Press Y= and enter $1.36x + 27.7$ for y_1.
Press GRAPH.



PRACTICE

In Exercises 1–5, refer to the table, which shows the total sales from men's clothing stores in the United States from 1997 to 2002.

Year	1997	1998	1999	2000	2001	2002
Sales (billions of dollars)	10.1	10.6	10.5	10.8	10.3	9.9

1. Make a scatter plot of the data. *Describe* the correlation.
2. Find the equation of the best-fitting line for the data.
3. Draw the best-fitting line for the data.

DRAW CONCLUSIONS

4. What does the value of r for the equation in Exercise 2 tell you about the correlation of the data?
5. **PREDICT** How could you use the best-fitting line to predict future sales of men's clothing? *Explain* your answer.