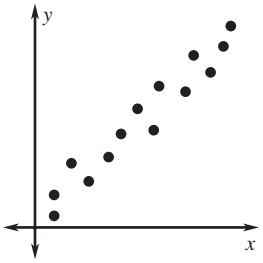
LESSON 29

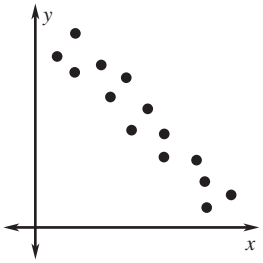
Scatter Plots

GOAL Make and interpret scatter plots; find best-fitting lines.

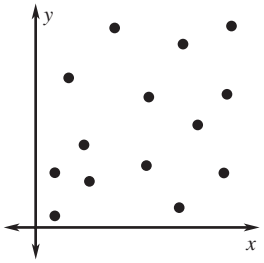
A **scatter plot** is a graph used to determine whether there is a relationship between two variables. You can make a scatter plot by representing paired, or bivariate, data as ordered pairs (x, y) and plotting them as points in a coordinate plane. Scatter plots show whether paired data have a *positive correlation*, a *negative correlation*, or *relatively no correlation*.



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



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



When x and y have no apparent relationship, the paired data have **relatively no correlation**.

EXAMPLE 1 Making a Scatter Plot

The table below gives the percent of women who read a weekday newspaper during several different years. Make a scatter plot of the data. Then tell whether the data have a *positive correlation,* a *negative correlation,* or *relatively no correlation.*

Year	1980	1982	1984	1986	1988	1990	1992	1994	1996
Percent of women	64.8	65.5	64.1	61.1	62.4	60.5	60.2	59.5	54.9

SOLUTION

Write the data as ordered pairs. Let x be the number of years since 1980. Let y be the percent of women who read a weekday newspaper.

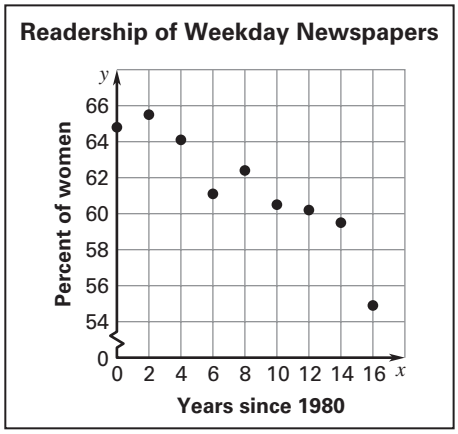
(0, 64.8), (2, 65.5), (4, 64.1),

(6, 61.1), (8, 62.4), (10, 60.5),

(12, 60.2), (14, 59.5), (16, 54.9)

Then plot the ordered pairs in a coordinate plane. The scatter plot is shown at the right.

Because y tends to decrease as x increases, the data have a negative correlation.



CHECK Example 1

1. Make a scatter plot of the data. Then tell whether the data have a *positive correlation*, a *negative correlation*, or *relatively no correlation*.

x	2	3	4	5	6	6	8	10
y	20	25	35	25	40	45	55	60

Best-Fitting Lines When data have a positive or negative correlation you can make predictions using a *best-fitting line*. A **best-fitting line** is a line that closely follows the pattern of a set of data points. In this lesson, you will visually approximate best-fitting lines.

EXAMPLE 2 Approximating a Best-Fitting Line

Use the data and the scatter plot from Example 1.

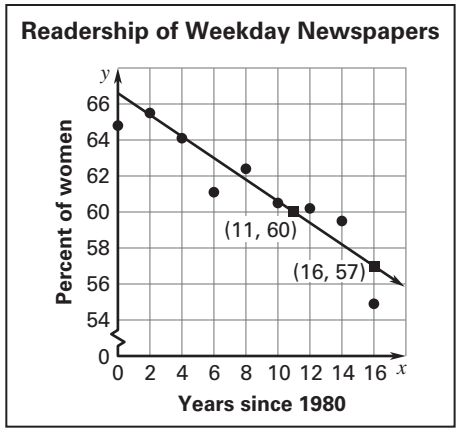
a. Sketch a line that appears to best fit the data.

b. Write an equation of the line.

c. Predict the percent of women who read a weekday newspaper in 2006.

SOLUTION

a. Sketch a line that closely follows the pattern of the data points. About half of the data points should be above the line and about half of the data points should be below the line.



b. The line passes through the points (11, 60) and (16, 57). So, the slope of the line is:

$$m = \frac{57 - 60}{16 - 11} = \frac{-3}{5} = -0.6$$

Use point-slope form to find an equation of the line.

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

$y - 60 = -0.6(x - 11)$ *Substitute for m, x_1, and y_1.*

$y = -0.6x + 66.6$ *Simplify.*

c. For the year 2006, the number of years since 1980 is $2006 - 1980 = 26$. To predict the percent of women who read a weekday newspaper in 2006, substitute 26 for x in the linear model.

$y = -0.6x + 66.6$ *Write linear model.*

$y = -0.6(26) + 66.6$ *Substitute 26 for x.*

$y = 51$ *Simplify.*

You can predict that 51% of women read a weekday newspaper in 2006.

CHECK Example 2

2. On the scatter plot you made in Exercise 1 on page 141, sketch a line that appears to best fit the data. Then write an equation of the line.

Activity

Using a Scatter Plot

1. For each of 20 CDs, find the number of songs on the CD and the total playing time to the nearest minute. Record the data in a table like the one shown.

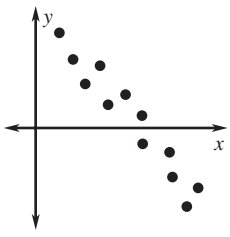
Songs	?	?	?
Total playing time	?	?	?

2. Make a scatter plot of the data. Let x represent the number of songs and let y represent the total playing time.
3. Describe the correlation between the number of songs on a CD and the total playing time of the CD.
4. Sketch a line that appears to best fit the data. Write an equation of the line.
5. Use your model to estimate the total playing time of a CD that has a different number of songs than the CDs you used.

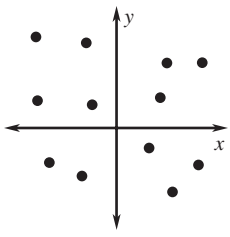
EXERCISES

Tell whether *x* and *y* have a *positive correlation,* a *negative correlation,* or *relatively no correlation.* Explain your reasoning.

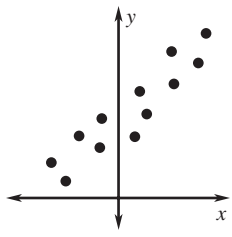
1.



2.



3.



Make a scatter plot of the data. Then tell whether the data have a *positive correlation,* a *negative correlation,* or *relatively no correlation.*

4.

x	−20	−16	−10	−5	−1	3	8	11	15	19
y	2	3	3	6	9	10	17	14	16	22

5.

x	−5	−2	0	1	4	5
y	3	1	0	−2	−5	−6

Make a scatter plot of the data. Sketch a line that appears to best fit the data. Then write an equation of the line.

6.

x	−5	−3	−2	−1.5	0	1	2	3	4	5
y	6	5	2	3.5	2	0.5	3	0	−1	1

7.

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

8.

x	1	1.5	2	3	4	5	6	7	7.5	8
y	7.5	6	6	5	4.5	5	3	3.5	4	3.5

In Exercises 9–11, use the table below, which gives a truck's maximum speed going uphill for various hill slopes.

Slope of hill (degrees)	2	5	8	10
Truck's maximum speed (mi/h)	76	58	35	20

9. Make a scatter plot of the data.

10. On your scatter plot sketch a line that appears to best fit the data. Then write an equation of the line.

11. The slope of a highway through the mountains is 12°. Use your model to predict the truck's maximum speed going uphill.

In Exercises 12–15, use the table below, which gives the number of paperback books bought at a book store and the total cost of the books.

Paperback books	1	2	3	4	5	5	6	9	10
Total cost	$8.50	$19.95	$24.50	$40.20	$42.50	$45.60	$50.35	$75.00	$80.05

12. Make a scatter plot of the data.

13. On your scatter plot sketch a line that appears to best fit the data. Then write an equation of the line.

14. Use your model to predict the total cost of buying 8 paperback books.

15. What does the slope of your best-fitting line represent in terms of the data? Explain.

16. Describe a real-life situation in which a collection of paired data might have the given relationship.

a. A positive correlation

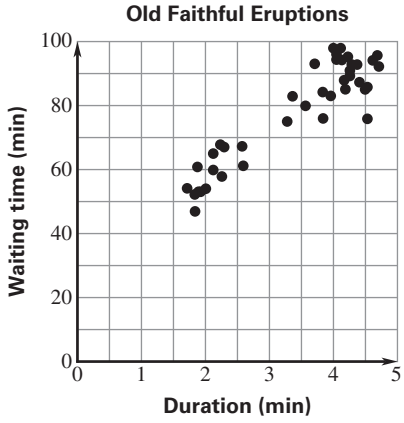
b. A negative correlation

c. Relatively no correlation

17. Tara makes a scatter plot of some data. She notices that as the y-values increase, the x-values decrease. What type of correlation do the data show? Explain.

In Exercises 18–20, use the following information.

The scatter plot displays data for Old Faithful, a geyser in Yellowstone National Park. The horizontal axis shows the duration of an eruption. The vertical axis shows the waiting time until the next eruption.

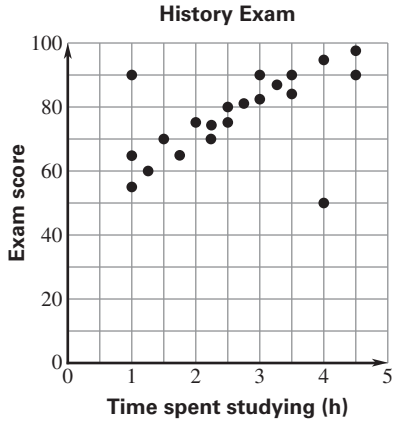


18. Describe any correlation you see in the scatter plot.

19. A *cluster* is a set of closely grouped data. Describe the clusters in the scatter plot.

20. What do the clusters tell you about eruptions of Old Faithful?

21. On a scatter plot, an *outlier* is a data point that is widely separated from the other data points.



a. Identify any outliers in the History Exam scatter plot by giving the coordinates of the points.

b. One student spent relatively little time studying but did very well on the exam. Which point represents this student's data?

c. Do the data show a correlation? If so, do you think the correlation is affected by the outliers? Explain.

d. A student not represented on the scatter plot studied for 2.5 hours. What do you think is a range of possible exam scores for which the student's data point would *not* be considered an outlier? Give your reasoning.

22. The spreadsheet shows the planets' mean distances from the Sun and the mean velocities at which the planets orbit the Sun.

	A	B	C
1	Planet	Mean distance from Sun (astronomical units)	Mean orbital velocity (km/sec)
2	Mercury	0.3871	47.89
3	Venus	0.7233	35.04
4	Earth	1.000	29.79
5	Mars	1.524	24.14
6	Jupiter	5.203	13.06
7	Saturn	9.539	9.64
8	Uranus	19.19	6.81
9	Neptune	30.06	5.43

a. Use a spreadsheet to make a scatter plot of the data.

b. Tell whether the data have a *positive correlation,* a *negative correlation,* or *relatively no correlation.*

c. What can you conclude about the planets based on your scatter plot?