

# Scatter Plots and Trend Lines

**Objectives** To write an equation of a trend line and of a line of best fit  
To use a trend line and a line of best fit to make predictions



Would you expect the number of albums downloaded to have an effect on CD sales?



## Getting Ready!

The table shows the number of digital albums downloaded per year and the number of CDs sold by manufacturers per year. What relationship exists between the two sets of data? Predict the number of CDs sold and the number of albums downloaded in 2010. Explain your reasoning.

Music Sales		
Year	Albums Downloaded (millions)	CDs Sold (millions)
2004	4.6	767
2005	13.6	705.4
2006	27.6	619.7
2007	42.5	511.1

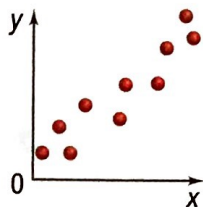
Source: Recording Industry Association of America

In the Solve It, the number of albums downloaded per year and the number of CDs sold per year are related.

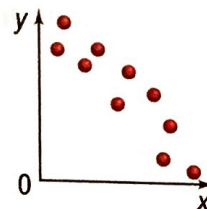
**Essential Understanding** You can determine whether two sets of numerical data are related by graphing them as ordered pairs. If the two sets of data are related, you may be able to use a line to estimate or predict values.

A **scatter plot** is a graph that relates two different sets of data by displaying them as ordered pairs. Most scatter plots are in the first quadrant of the coordinate plane because the data are usually positive numbers.

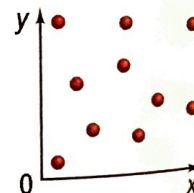
You can use scatter plots to find trends in data. The scatter plots below show the three types of relationships that two sets of data may have.



When  $y$  tends to increase as  $x$  increases, the two sets of data have a **positive correlation**.



When  $y$  tends to decrease as  $x$  increases, the two sets of data have a **negative correlation**.



When  $x$  and  $y$  are not related, the two sets of data have **no correlation**.



### Lesson Vocabulary

- scatter plot
- positive correlation
- negative correlation
- no correlation
- trend line
- interpolation
- extrapolation
- line of best fit
- correlation coefficient
- causation

## Problem 1 Making a Scatter Plot and Describing Its Correlation

**Temperature** The table shows the altitude of an airplane and the temperature outside the plane.

Altitude (m)	0	500	1000	1500	2000	2500	3000	3500	4000	4500	5000
Temperature (°F)	59.0	59.2	61.3	55.5	41.6	29.8	29.9	18.1	26.2	12.4	0.6

### Think

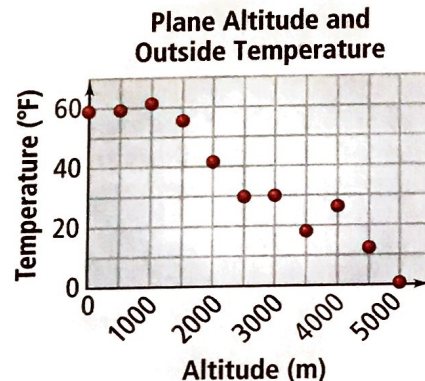
The highest altitude is 5000 m. So a reasonable scale on the altitude axis is 0 to 5500 with every 1000 m labeled. You can use similar reasoning to label the temperature axis.

**A** Make a scatter plot of the data.

Treat the data as ordered pairs. For the altitude of 1500 m and the temperature of 55.5°F, plot (1500, 55.5).

**B** What type of relationship does the scatter plot show?

The temperature outside the plane tends to decrease as the altitude of the plane increases. So the data have a negative correlation.



**Got It?** 1. a. Make a scatter plot of the data in the table below. What type of relationship does the scatter plot show?

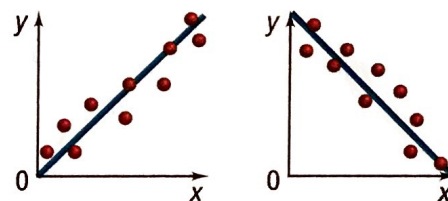
Dollars Spent	10	11	9	10	13	5	8	4
Gallons Bought	2.5	2.8	2.3	2.6	3.3	1.3	2.2	1.1

**b. Reasoning** Consider the population of a city and the number of letters in the name of the city. Would you expect a *positive correlation*, a *negative correlation*, or *no correlation* between the two sets of data? Explain your reasoning.

When two sets of data have a positive or negative correlation, you can use a trend line to show the correlation more clearly. A **trend line** is a line on a scatter plot, drawn near the points, that shows a correlation.

You can use a trend line to estimate a value between two known data values or to predict a value outside the range of known data values.

**Interpolation** is estimating a value between two known values. **Extrapolation** is predicting a value outside the range of known values.



## Problem 2 Writing an Equation of a Trend Line STEM

**Biology** Make a scatter plot of the data at the right. What is the approximate weight of a 7-month-old panda?

**Step 1** Make a scatter plot and draw a trend line. Estimate the coordinates of two points on the line.

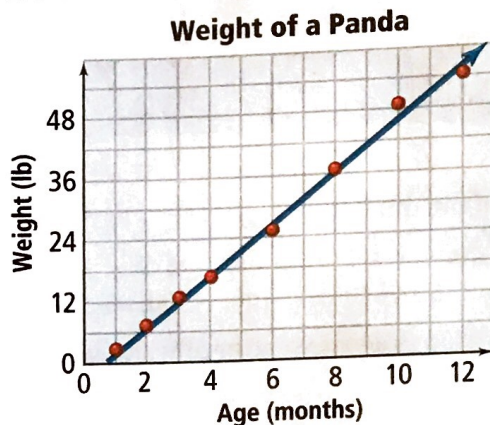
Weight of a Panda

Age (months)	Weight (lb)
1	2.5
2	7.6
3	12.5
4	17.1
6	24.3
8	37.9
10	49.2
12	54.9

### Plan

**How do you draw an accurate trend line?**

An accurate trend line should fit the data closely. There should be about the same number of points above the line as below it.



Two points on the trend line are (4, 17.1) and (8, 37.9).

**Step 2** Write an equation of the trend line.

$$m = \frac{y_2 - y_1}{x_2 - x_1} = \frac{37.9 - 17.1}{8 - 4} = \frac{20.8}{4} = 5.2$$

$$y - y_1 = m(x - x_1)$$

$$y - 17.1 = 5.2(x - 4)$$

$$y - 17.1 = 5.2x - 20.8$$

$$y = 5.2x - 3.7$$

Find the slope of the trend line.

Use point-slope form.

Substitute 5.2 for  $m$  and (4, 17.1) for  $(x_1, y_1)$ .

Distributive Property

Add 17.1 to each side.

**Step 3** Estimate the weight of a 7-month-old panda.

$$y = 5.2(7) - 3.7$$

$$y = 32.7$$

Substitute 7 for  $x$ .

Simplify.

The weight of a 7-month-old panda is about 32.7 lb.

### Think

**How can you check the reasonableness of your answer?**

Since  $x = 7$  is visible on the graph, find its corresponding  $y$ -value. When  $x = 7$ ,  $y \approx 32.7$ . So the estimate is reasonable.



**Got It?** 2. a. Make a scatter plot of the data below. Draw a trend line and write its equation. What is the approximate body length of a 7-month-old panda?

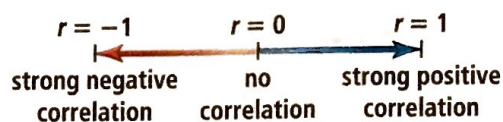
Body Length of a Panda

Age (months)	1	2	3	4	5	6	8	9
Body Length (in.)	8.0	11.75	15.5	16.7	20.1	22.2	26.5	29.0

b. **Reasoning** Do you think you can use your model to extrapolate the body length of a 3-year-old panda? Explain.

The trend line that shows the relationship between two sets of data most accurately is called the **line of best fit**. A graphing calculator computes the equation of the line of best fit using a method called linear regression.

The graphing calculator also gives you the **correlation coefficient**  $r$ , a number from  $-1$  to  $1$ , that tells you how closely the equation models the data.



The nearer  $r$  is to  $1$  or  $-1$ , the more closely the data cluster around the line of best fit. If  $r$  is near  $1$ , the data lie close to a line of best fit with positive slope. If  $r$  is near  $-1$ , the data lie close to a line of best fit with negative slope.

### Problem 3 Finding the Line of Best Fit

**College Tuition** Use a graphing calculator to find the equation of the line of best fit for the data at the right. What is the correlation coefficient to three decimal places? Predict the cost of attending in the 2012–2013 academic year.

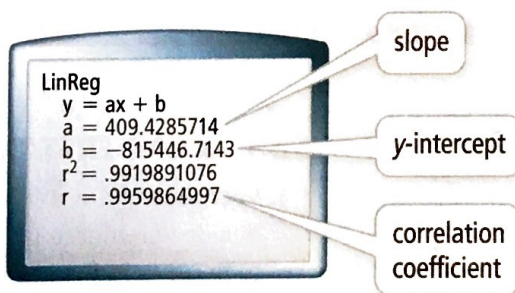
**Average Tuition and Fees at Public 4-Year Colleges**

Academic Year	Cost (\$)
2000–2001	3508
2001–2002	3766
2002–2003	4098
2003–2004	4645
2004–2005	5126
2005–2006	5492
2006–2007	5836

Source: The College Board

**Step 1** Press **stat**. From the **EDIT** menu, choose **Edit**. Enter the years into  $L_1$ . Let  $x = 2000$  represent academic year 2000–2001,  $x = 2001$  represent 2001–2002, and so on. Enter the costs into  $L_2$ .

**Step 2** Press **stat**. Choose **LinReg(ax + b)** from the **CALC** menu. Press **enter** to find the equation of the line of best fit and the correlation coefficient. The calculator uses the form  $y = ax + b$  for the equation.



#### Think

**What does the value of the correlation coefficient mean?**

The correlation coefficient of 0.996 is close to 1. So there is a strong positive correlation between the academic year and the cost of attending college.

Round to the nearest hundredth. The equation of the line of best fit is  $y = 409.43x - 815,446.71$ . The correlation coefficient is about 0.996.



**Step 3** Predict the cost of attending in the 2012–2013 academic year.

$$y = 409.43x - 815,446.71 \quad \text{Use the equation of the line of best fit.}$$

$$y = 409.43(2012) - 815,446.71 \quad \text{Substitute 2012 for } x.$$

$$y \approx 8326 \quad \text{Simplify. Round to the nearest whole number.}$$

The cost of attending a four-year public college in the 2012–2013 academic year is predicted to be about \$8326.

-   **Got It?** 3. a. Predict the cost of attending in the 2016–2017 academic year.  
b. **Reasoning** What does the slope of the line of best fit in Problem 3 tell you about the rate of change in the cost?

**Causation** is when a change in one quantity causes a change in a second quantity. A correlation between quantities does not always imply causation.

### **Problem 4** Identifying Whether Relationships Are Causal


In the following situations, is there likely to be a correlation? If so, does the correlation reflect a causal relationship? Explain.

**A** the number of loaves of bread baked and the amount of flour used

There is a positive correlation and also a causal relationship. As the number of loaves of bread baked increases, the amount of flour used increases.

**B** the number of mailboxes and the number of firefighters in a city

There is likely to be a positive correlation because both the number of mailboxes and the number of firefighters tend to increase as the population of a city increases. However, installing more mailboxes will not *cause* the number of firefighters to increase, so there is no causal relationship.

-  **Got It?** 4. In the following situations, is there likely to be a correlation? If so, does the correlation reflect a causal relationship? Explain.
- the cost of a family's vacation and the size of their house
  - the time spent exercising and the number of Calories burned

### Think

Causal relationships always have a correlation. However, two data sets that have a correlation may not have a causal relationship.

## Lesson Check

### Do you know HOW?

Use the table.




Average Maximum Daily Temperature in January for Northern Latitudes

Latitude ( $^{\circ}$ N)	35	33	30	25	43	40	39
Temperature ( $^{\circ}$ F)	46	52	67	76	32	37	44

SOURCE: U.S. Department of Commerce

- Make a scatter plot of the data. What type of relationship does the scatter plot show?
- Draw a trend line and write its equation.
- Predict the average maximum daily temperature in January at a latitude of  $50^{\circ}$  N.

### Do you UNDERSTAND? MATHEMATICAL PRACTICES

-  4. **Vocabulary** Given a set of data pairs, how would you decide whether to use interpolation or extrapolation to find a certain value?
-  5. **Compare and Contrast** How are a trend line and the line of best fit for a set of data pairs similar? How are they different?
-  6. **Error Analysis** Refer to the table below. A student says that the data have a negative correlation because as  $x$  decreases,  $y$  also decreases. What is the student's error?

$x$	10	7	5	4	1	0
$y$	1	0	-2	-4	-7	-9

## A Practice

For each table, make a scatter plot of the data. Describe the type of correlation the scatter plot shows.

← See Problem 1.

7.

Jeans Sales				
Average Price (\$)	21	28	36	40
Number Sold	130	112	82	65

8.

Gasoline Purchases					
Dollars Spent	10	11	9	8	13
Gallons Bought	2.6	3	2.4	2.2	3.5

**Theme Parks** Use the table below for Exercises 9 and 10.

← See Problem 2.

Attendance and Revenue at U.S. Theme Parks									
Year	1990	1992	1994	1996	1998	2000	2002	2004	2006
Attendance (millions)	253	267	267	290	300	317	324	328	335
Revenue (billions of dollars)	5.7	6.5	7.0	7.9	8.7	9.6	9.9	10.8	11.5

SOURCE: International Association of Amusement Parks and Attractions

- Make a scatter plot of the data pairs (year, attendance). Draw a trend line and write its equation. Estimate the attendance at U.S. theme parks in 2005.
- Make a scatter plot of the data pairs (year, revenue). Draw a trend line and write its equation. Predict the revenue at U.S. theme parks in 2012.
- Entertainment** Use a graphing calculator to find the equation of the line of best fit for the data in the table. Find the value of the correlation coefficient  $r$  to three decimal places. Then predict the number of movie tickets sold in the U.S. in 2014.

← See Problem 3.

Movie Tickets Sold in U.S. by Year										
Year	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
Tickets Sold (millions)	1289	1311	1340	1339	1406	1421	1470	1415	1472	1470

SOURCE: Motion Picture Association of America

In each situation, tell whether a correlation is likely. If it is, tell whether the correlation reflects a causal relationship. Explain your reasoning.

← See Problem 4.

- the amount of time you study for a test and the score you receive
- a person's height and the number of letters in the person's name
- the shoe size and the salary of a teacher
- the price of hamburger at a grocery store and the amount of hamburger sold

## B Apply

- Open-Ended** Describe three real-world situations: one with a positive correlation, one with a negative correlation, and one with no correlation.

- © 17. **Writing** Give two data sets that are correlated but do *not* have a causal relationship.
18. **Business** During one month at a local deli, the amount of ham sold decreased as the amount of turkey sold increased. Is this an example of *positive correlation*, *negative correlation*, or *no correlation*?
- © 19. **Think About a Plan** Students measured the diameters and circumferences of the tops of a variety of cylinders. Below are the data that they collected. Estimate the diameter of a cylinder with circumference 22 cm.

Cylinder Tops										
Diameter (cm)	3	3	5	6	8	8	9.5	10	10	12
Circumference (cm)	9.3	9.5	16	18.8	25	25.6	29.5	31.5	30.9	39.5


- How can you use a scatter plot to find an equation of a trend line?
- How can you use the equation of the trend line to make an estimate?

- © 20. **U.S. Population** Use the data below.

Estimated Population of the United States (thousands)							
Year	2000	2001	2002	2003	2004	2005	2006
Male	138,482	140,079	141,592	142,937	144,467	145,973	147,512
Female	143,734	145,147	146,533	147,858	149,170	150,533	151,886

SOURCE: U.S. Census Bureau

- Make a scatter plot of the data pairs (male population, female population).
- Draw a trend line and write its equation.
- Use your equation to predict the U.S. female population if the U.S. male population increases to 150,000,000.
- Reasoning** Consider a scatter plot of the data pairs (year, male population). Would it be reasonable to use this scatter plot to predict the U.S. male population in 2035? Explain your reasoning.

-  21. a. **Graphing Calculator** Use a graphing calculator to find the equation of the line of best fit for the data below. Let  $x = 8$  represent 1998,  $x = 9$  represent 1999, and so on.

U.S. Computer and Video Game Unit Sales										
Year	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
Unit Sales (millions)	152.4	184.5	196.3	210.3	225.8	240.9	249.5	229.5	241.6	267.9

SOURCE: The NPD Group/Retail Tracking Service

- What is the slope of the line of best fit? What does the slope mean in terms of the number of computer and video game units sold?
- What is the  $y$ -intercept of the line of best fit? What does the  $y$ -intercept mean in terms of the number of computer and video game units sold?

**Challenge**

22. a. Make a scatter plot of the data below. Then find the equation of the line of best fit. Draw the line of best fit on your scatter plot.

Car Stopping Distances								
Speed (mi/h)	10	15	20	25	30	35	40	45
Stopping Distance (ft)	27	44	63	85	109	136	164	196

- b. Use your equation to predict the stopping distance at 90 mi/h.
- c. **Reasoning** The actual stopping distance at 90 mi/h is close to 584 ft. Why do you think this distance is not close to your prediction?
- d. Suppose you plot (90, 584) on your scatter plot. What effect would it have on the slope and y-intercept of the line of best fit you found in part (a)?

**Standardized Test Prep**

SAT/ACT

23. Suppose you survey each school in your state. What relationship would you expect between the number of students and the number of teachers in each school?

- (A) positive correlation                      (C) no correlation  
 (B) negative correlation                      (D) none of the above

24. A horizontal line passes through (5, -2). Which other point is also on the line?

- (F) (5, 2)                      (G) (-5, -2)                      (H) (-5, 2)                      (I) (5, 0)

25. When 18 gal of water are pumped into an empty tank, the tank is filled to three fourths of its capacity. How many gallons of water does the tank hold?

- (A) 12                      (B) 13.5                      (C) 18.5                      (D) 24

Short Response

26. The table shows the balance of a student's bank account at various times. Estimate how much money is in the student's bank account in Week 6. Justify your answer.

Weekly Account Balance					
Week	1	3	4	7	9
Account Balance	\$35	\$68	\$85	\$105	\$136

**Mixed Review**

Write an equation of the line in slope-intercept form that passes through the given point and is parallel to the graph of the given equation.

◀ See Lesson 5-6.

27.  $y = 5x + 1; (2, -3)$

28.  $y = -x - 9; (0, 5)$

29.  $2x + 3y = 9; (-1, 4)$

**Get Ready!** To prepare for Lesson 5-8, do Exercises 30-33.

Find each absolute value.

◀ See Lesson 1-5.

30.  $|2 - 7|$

31.  $|7 - 7|$

32.  $|56 - 38|$

33.  $|-24 + 12|$