Modeling Real-World Data with Linear Functions

OBJECTIVES
- Draw and analyze scatter plots.
- Write a prediction equation and draw best-fit lines.
- Use a graphing calculator to compute correlation coefficients to determine goodness of fit.
- Solve problems using prediction equation models.

Real World Application

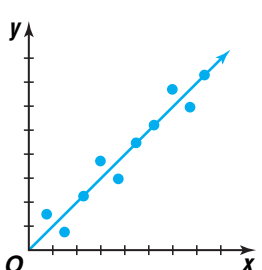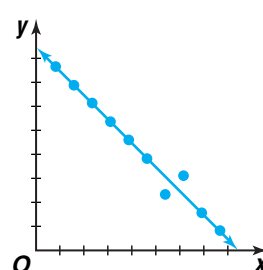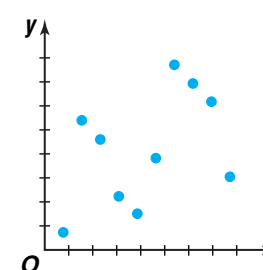**Education** The cost of attending college is steadily increasing. However, it can be a good investment since on average, the higher your level of education, the greater your earning potential. The chart shows the average tuition and fees for a full-time resident student at a public four-year college. Estimate the average college cost in the academic year beginning in 2006 if tuition and fees continue at this rate. *This problem will be solved in Example 1.*

Academic Year	Tuition and Fees
1990–1991	2159
1991–1992	2410
1992–1993	2349
1993–1994	2537
1994–1995	2681
1995–1996	2811
1996–1997	2975
1997–1998	3111
1998–1999	3243

Source: The College Board and National Center for Educational Statistics

As you look at the college tuition costs, it is difficult to visualize how quickly the costs are increasing. When real-life data is collected, the data graphed usually does not form a perfectly straight line. However, the graph may approximate a linear relationship. When this is the case, a **best-fit line** can be drawn, and a **prediction equation** that models the data can be determined. Study the **scatter plots** below.

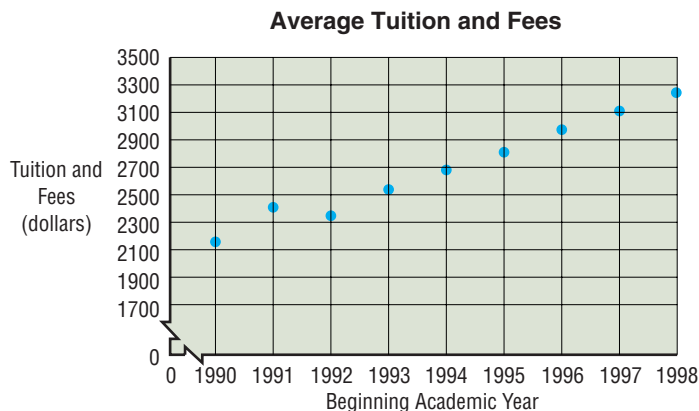| Linear Relationship		No Pattern
This scatter plot suggests a linear relationship. Notice that many of the points lie on a line, with the rest very close to it. Since the line has a positive slope, these data have a positive relationship.	This scatter plot also implies a linear relationship. However, the slope of the line suggested by the data is negative.	The points in this scatter plot are very dispersed and do not appear to form a linear pattern.

A prediction equation can be determined using a process similar to determining the equation of a line using two points. The process is dependent upon your judgment. You decide which two points on the line are used to find the slope and intercept. Your prediction equation may be different from someone else's. A prediction equation is used when a rough estimate is sufficient.

Example

1 **EDUCATION** **Refer to the application at the beginning of the lesson. Predict the average college cost in the academic year beginning in 2006.**

Graph the data. Use the starting year as the independent variable and the tuition and fees as the dependent variable.

Average Tuition and Fees

Tuition and Fees (dollars) — vertical axis: 1700, 1900, 2100, 2300, 2500, 2700, 2900, 3100, 3300, 3500
Beginning Academic Year — horizontal axis: 1990, 1991, 1992, 1993, 1994, 1995, 1996, 1997, 1998

Select two points that appear to represent the data. We chose (1992, 2349) and (1997, 3111). Determine the slope of the line.

$$m = \frac{y_2 - y_1}{x_2 - x_1} \qquad \textit{Definition of slope}$$

$$= \frac{3111 - 2349}{1997 - 1992} \qquad \textit{(x}_1\textit{, y}_1\textit{) = (1992, 2349), (x}_2\textit{, y}_2\textit{) = (1997, 3111)}$$

$$= \frac{762}{5} \text{ or } 152.4$$

Now use one of the ordered pairs, such as (1992, 2349), and the slope in the point-slope form of the equation.

$$y - y_1 = m(x - x_1) \qquad \textit{Point-slope form of an equation}$$

$$y - 2349 = 152.4(x - 1992) \qquad \textit{(x}_1\textit{, y}_1\textit{) = (1992, 2349), and m = 152.4}$$

$$y = 152.4x - 301{,}231.8$$

Thus, a prediction equation is $y = 152.4x - 301{,}231.8$. Substitute 2006 for x to estimate the average tuition and fees for the year 2006.

$$y = 152.4x - 301{,}231.8$$

$$y = 152.4(2006) - 301{,}231.8$$

$$y = 4482.6$$

According to this prediction equation, the average tuition and fees will be $4482.60 in the academic year beginning in 2006. *Use a different pair of points to find another prediction equation. How does it compare with this one?*

CONTENTS

Data that are linear in nature will have varying degrees of **goodness of fit** to the lines of fit. Various formulas are often used to find a **correlation coefficient** that describes the nature of the data. The more closely the data fit a line, the closer the correlation coefficient r approaches 1 or -1. Positive correlation coefficients are associated with linear data having positive slopes, and negative correlation coefficients are associated with negative slopes. Thus, the more linear the data, the more closely the correlation coefficient approaches 1 or -1.

$0 < r \leq 0.5$
positive and weak

$0.5 \leq r \leq 0.75$
moderately positive

$0.75 \leq r \leq 1$
strongly positive

$-0.5 \leq r < 0$

$-0.75 \leq r \leq -0.5$

$-1 \leq r \leq -0.75$
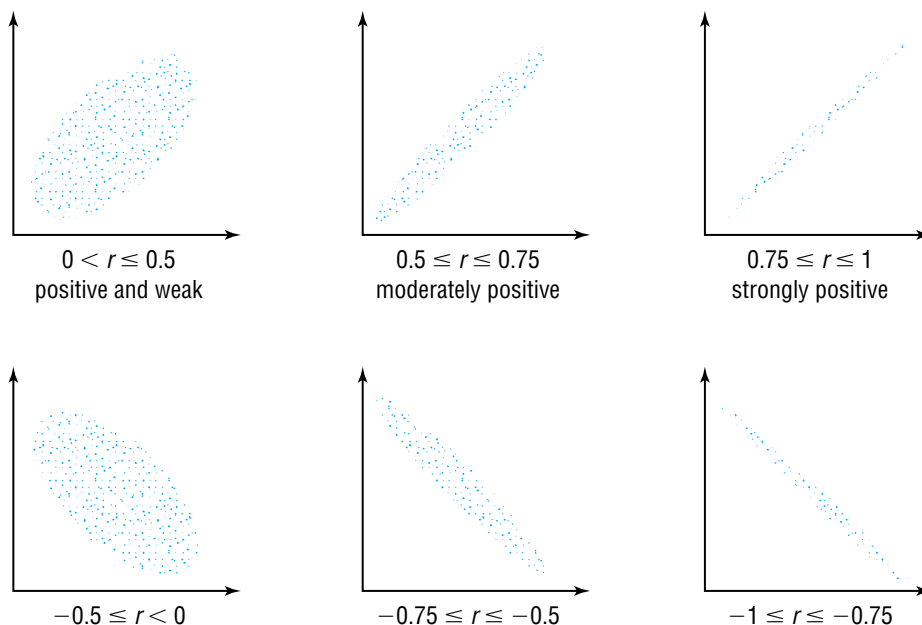
Statisticians normally use precise procedures, often relying on computers to determine correlation coefficients. The graphing calculator uses the **Pearson product-moment correlation,** which is represented by r. When using these methods, the best fit-line is often called a **regression line.**

Example

Real World Application

2 **NUTRITION**   The table contains the fat grams and Calories in various fast-food chicken sandwiches.

a. Use a graphing calculator to find the equation of the regression line and the Pearson product-moment correlation.

b. Use the equation to predict the number of Calories in a chicken sandwich that has 20 grams of fat.

Chicken Sandwich (cooking method)	Fat (grams)	Calories
A (breaded)	28	536
B (grilled)	20	430
C (chicken salad)	33	680
D (broiled)	29	550
E (breaded)	43	710
F (grilled)	12	390
G (breaded)	9	300
H (chicken salad)	5	320
I (breaded)	26	530
J (breaded)	18	440
K (grilled)	8	310

CONTENTS

a. Enter the data for fat grams in list L1 and the data for Calories in list L2. Draw a scatter plot relating the fat grams, *x,* and the Calories, *y.*

Then use the linear regression statistics to find the equation of the regression line and the correlation coefficient.

The Pearson product-moment correlation is about 0.98. The correlation between grams of fat and Calories is strongly positive. Because of the strong relationship, the equation of the regression line can be used to make predictions.

b. When rounding to the nearest tenth, the equation of the regression line is $y = 11.6x + 228.3$. Thus, there are about $y = 11.6(20) + 228.3$ or 460.3 Calories in a chicken sandwich with 20 grams of fat.

[0, 45] scl: 1 by [250, 750] scl: 50

```
LinReg
 y=ax+b
 a=11.62131837
 b=228.3159505
 r²=.9539063457
 r=.9766812918
```
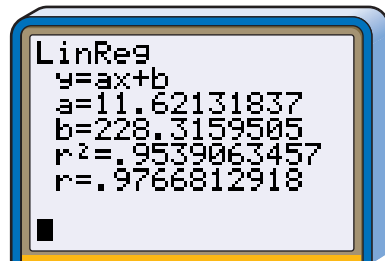
It should be noted that even when there is a large correlation coefficient, you cannot assume that there is a "cause and effect" relationship between the two related variables.

CHECK FOR UNDERSTANDING

Communicating Mathematics

Read and study the lesson to answer each question.

1. **Explain** what the slope in a best-fit line represents.

2. **Describe** three different methods for finding a best-fit line for a set of data.

3. **Write** about a set of real-world data that you think would show a negative correlation.

Guided Practice

Complete parts a–d for each set of data given in Exercises 4 and 5.
a. **Graph the data on a scatter plot.**
b. **Use two ordered pairs to write the equation of a best-fit line.**
c. **Use a graphing calculator to find an equation of the regression line for the data. What is the correlation coefficient?**
d. **If the equation of the regression line shows a moderate or strong relationship, predict the missing value. Explain whether you think the prediction is reliable.**

4. **Economics** The table shows the average amount that an American spent on durable goods in several years.

Personal Consumption Expenditures for Durable Goods									
Year	1990	1991	1992	1993	1994	1995	1996	1997	2010
Personal Consumption ($)	1910	1800	1881	2083	2266	2305	2389	2461	?

Source: U.S. Dept. of Commerce

Lesson 1-6 Modeling Real-World Data with Linear Functions **41**

CONTENTS

5. Education Do you share a computer at school? The table shows the average number of students per computer in public schools in the United States.

Students per Computer								
Academic Year	1983–1984	1984–1985	1985–1986	1986–1987	1987–1988	1988–1989	1989–1990	1990–1991
Average	125	75	50	37	32	25	22	20

Academic Year	1991–1992	1992–1993	1993–1994	1994–1995	1995–1996	1996–1997	?
Average	18	16	14	10.5	10	7.8	1

Source: QED's Technology in Public Schools

EXERCISES

Complete parts a–d for each set of data given in Exercises 6–11.

a. **Graph the data on a scatter plot.**

b. **Use two ordered pairs to write the equation of a best-fit line.**

c. **Use a graphing calculator to find an equation of the regression line for the data. What is the correlation coefficient?**

d. **If the equation of the regression line shows a moderate or strong relationship, predict the missing value. Explain whether you think the prediction is reliable.**

6. Sports The table shows the number of years coaching and the number of wins as of the end of the 1999 season for selected professional football coaches.

NFL Coach	Years	Wins
Don Shula	33	347
George Halas	40	324
Tom Landry	29	270
Curly Lambeau	33	229
Chuck Noll	23	209
Chuck Knox	22	193
Dan Reeves	19	177
Paul Brown	21	170
Bud Grant	18	168
Steve Owen	23	153
Marv Levy	17	?

Source: World Almanac

7. Economics Per capita personal income is the average personal income for a nation. The table shows the per capita personal income for the United States for several years.

Year	1990	1991	1992	1993	1994	1995	1996	1997	2005
Personal Income ($)	18,477	19,100	19,802	20,810	21,846	23,233	24,457	25,660	?

Source: U.S. Dept. of Commerce

www.amc.glencoe.com/self_check_quiz

CONTENTS

8. Transportation   Do you think the weight of a car is related to its fuel economy? The table shows the weight in hundreds of pounds and the average miles per gallon for selected 1999 cars.

Weight (100 pounds)	17.5	20.0	22.5	22.5	22.5	25.0	27.5	35.0	45.0
Fuel Economy (mpg)	65.4	49.0	59.2	41.1	38.9	40.7	46.9	27.7	?

Source: U.S. Environmental Protection Agency

9. Botany   Acorns were one of the most important foods of the Native Americans. They pulverized the acorns, extracted the bitter taste, and then cooked them in various ways. The table shows the size of acorns and the geographic area covered by different species of oak.

Acorn size (cm^3)	0.3	0.9	1.1	2.0	3.4	4.8	8.1	10.5	17.1
Range (100 km^2)	233	7985	10,161	17,042	7900	3978	28,389	7646	?

Source: Journal of Biogeography

10. Employment   Women have changed their role in American society in recent decades. The table shows the percent of working women who hold managerial or professional jobs.

Percent of Working Women in Managerial or Professional Occupations									
Year	1986	1988	1990	1992	1993	1994	1995	1996	1997	2008
Percent	23.7	25.2	26.2	27.4	28.3	28.7	29.4	30.3	30.8	?

Source: U.S. Dept. of Labor

11. Demographics   The world's population is growing at a rapid rate. The table shows the number of millions of people on Earth at different years.

World Population							
Year	1	1650	1850	1930	1975	1998	2010
Population (millions)	200	500	1000	2000	4000	5900	?

Source: World Almanac

12. Critical Thinking   Different correlation coefficients are acceptable for different situations. For each situation, give a specific example and explain your reasoning.

a. When would a correlation coefficient of less than 0.99 be considered unsatisfactory?

b. When would a correlation coefficient of 0.6 be considered good?

c. When would a strong negative correlation coefficient be desirable?

13. Critical Thinking The table shows the median salaries of American men and women for several years. According to the data, will the women's median salary ever be equal to the men's? If so, predict the year. Explain.

Median Salary ($)					
Year	Men's	Women's	Year	Men's	Women's
1985	16,311	7217	1991	20,469	10,476
1986	17,114	7610	1992	20,455	10,714
1987	17,786	8295	1993	21,102	11,046
1988	18,908	8884	1994	21,720	11,466
1989	19,893	9624	1995	22,562	12,130
1990	20,293	10,070	1996	23,834	12,815

Source: U.S. Bureau of the Census

Mixed Review

14. Business During the month of January, Fransworth Computer Center sold 24 computers of a certain model and 40 companion printers. The total sales on these two items for the month of January was $38,736. In February, they sold 30 of the computers and 50 printers. *(Lesson 1-5)*

a. Assuming the prices stayed constant during the months of January and February, is it possible that their February sales could have totaled $51,470 on these two items? Explain.

b. Assuming the prices stayed constant during the months of January and February, is it possible that their February sales could have totaled $48,420 on these two items? Explain.

15. Line ℓ passes through $A(-3, -4)$ and has a slope of -6. What is the standard form of the equation for line ℓ? *(Lesson 1-4)*

16. Economics The equation $y = 0.82x + 24$, where $x \geq 0$, models a relationship between a nation's disposable income, x in billions of dollars, and personal consumption expenditures, y in billions of dollars. Economists call this type of equation a *consumption function*. *(Lesson 1-3)*

a. Graph the consumption function.

b. Name the y-intercept.

c. Explain the significance of the y-intercept and the slope.

17. Find $[f \circ g](x)$ and $[g \circ f](x)$ if $f(x) = x^3$ and $g(x) = x + 1$. *(Lesson 1-2)*

18. Determine if the relation $\{(2, 4), (4, 2), (-2, 4), (-4, 2)\}$ is a function. Explain. *(Lesson 1-1)*

19. SAT/ACT Practice Choose the equation that is represented by the graph.

A $y = 3x - 1$

B $y = \frac{1}{3}x - 1$

C $y = 1 - 3x$

D $y = 1 - \frac{1}{3}x$

E none of these

CONTENTS