ALGEBRA

Exploring Linear Relations

GAIL F. BURRILL AND PATRICK HOPFENSPERGER

DATA-DRIVEN MATHEMATICS

DALE SEYMOUR PUBLICATIONS®

Exploring Linear Relations

DATA-DRIVEN MATHEMATICS

Gail F. Burrill and Patrick Hopfensperger

Dale Seymour Publications®

This material was produced as a part of the American Statistical Association's Project "A Data-Driven Curriculum Strand for High School" with funding through the National Science Foundation, Grant #MDR-9054648. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Managing Editor: Cathy Anderson

Senior Math Editor: Carol Zacny

Project Editor: Mary Myers

Production Manager: Janet Yearian

Senior Production Coordinator: Alan Noyes

Design Manager: Jeff Kelly

Text Design: Christy Butterfield

This book is published by Dale Seymour Publications®, an imprint of the Alternative Publishing Group of Addison Wesley Longman, Inc.

Printed in the United States of America.

Order number DS21161

ISBN 1-57232-210-1

1 2 3 4 5 6 7 8 9 10-ML-02 01 00 99 98 97

This Book Is Printed On Recycled Paper

DALE SEYMOUR PUBLICATIONS®

Authors

Gail F. Burrill
Whitnall High School
Greenfield, Wisconsin
University of Wisconsin-Madison
Madison, Wisconsin

Patrick Hopfensperger
Homestead High School
Mequon, Wisconsin

Data-Driven Mathematics Leadership Team

Miriam Clifford
Nicolet High School
Glendale, Wisconsin

James M. Landwehr
Bell Laboratories
Lucent Technologies
Murray Hill, New Jersey

Kenneth Sherrick
Berlin High School
Berlin, Connecticut

Gail F. Burrill
Whitnall High School
Greenfield, Wisconsin
University of Wisconsin-Madison
Madison, Wisconsin

Richard Scheaffer
University of Florida
Gainesville, Florida

Consultants

Jack Burrill
Whitnall High School
Greenfield, Wisconsin
University of Wisconsin-Madison
Madison, Wisconsin

Vince O'Connor
Milwaukee Public Schools
Milwaukee, Wisconsin

Emily Errthum
Homestead High School
Mequon, Wisconsin

Maria Mastromatteo
Brown Middle School
Ravenna, Ohio

Henry Kranendonk
Rufus King High School
Milwaukee, Wisconsin

Jeffrey Witmer
Oberlin College
Oberlin, Ohio

Acknowledgments

The authors would like to thank the following people for their assistance during the preparation of this module:

- The many teachers who reviewed drafts and participated in the field tests of the manuscripts
- The members of the *Data-Driven Mathematics* leadership team, the consultants and writers
- Nancy Kinard, Ron Moreland, Peggy Layton, and Kay Williams for their advice and suggestions in the early stages of the writing
- Jack Price and Ann Watkins for their thoughtful and careful review of the early drafts
- Kathryn Rowe and Wayne Jones for their help in organizing the field test process and the Leadership Workshops
- Barbara Shannon for many hours of word processing and secretarial services
- Jean Moon for her advice on how to improve the field test process
- Beth and Bryan Cole for writing answers for the Teacher's Edition
- The many students from Homestead High School and Whitnall High School who helped shape the ideas as they were being developed

Table of Contents

About *Data-Driven Mathematics*

Historically, the purposes of secondary-school mathematics have been to provide students with opportunities to acquire the mathematical knowledge needed for daily life and effective citizenship, to prepare students for the workforce, and to prepare students for postsecondary education. In order to accomplish these purposes today, students must be able to analyze, interpret, and communicate information from data.

Data-Driven Mathematics is a series of modules meant to complement a mathematics curriculum in the process of reform. The modules offer materials that integrate data analysis with high-school mathematics courses. Using these materials will help teachers motivate, develop, and reinforce concepts taught in current texts. The materials incorporate major concepts from data analysis to provide realistic situations for the development of mathematical knowledge and realistic opportunities for practice. The extensive use of real data provides opportunities for students to engage in meaningful mathematics. The use of real-world examples increases student motivation and provides opportunities to apply the mathematics taught in secondary school.

The project, funded by the National Science Foundation, included writing and field testing the modules, and holding conferences for teachers to introduce them to the materials and to seek their input on the form and direction of the modules. The modules are the result of a collaboration between statisticians and teachers who have agreed on statistical concepts most important for students to know and the relationship of these concepts to the secondary mathematics curriculum.

A diagram of the modules and possible relationships to the curriculum is on the back cover of each module.

Using This Module

Studying mathematics involves studying the relationship between two variables. Numerous examples of these relationships are found in the world outside of the classroom, such as the cost of living over time, the speed of a car and the rate of fuel consumption, the number of movie theaters showing a given movie, and the box-office revenue for that movie. In this module, you will explore real-world relationships as you make and interpret scatter plots, looking for patterns that you can describe mathematically. Most of the relationships you will study in the module are constant, and their graphs form straight lines. These relationships are called *linear relations.* The study of *linearity,* or linear relations, involves analyzing the rate of change between two variables, drawing a line that summarizes the relationships, and finding an equation that describes the line. You will use tables, graphs, and equations to estimate values and predict what might happen in the future as you pose answers to questions such as "Will women's income ever catch up with men's income?" and "How much will a new Corvette cost in 1998?"

Understanding the assumptions involved in making such predictions is an important part of the module—so is being able to describe the consistency of a relationship and the accuracy of using a particular line to summarize the relationship. When you have finished the module, you will have thoroughly investigated linear relations and will also understand how mathematics is used in the world around you.

Content

Algebra: You will be able to

- Find and interpret slope as a rate of change.
- Write the equation of a line from given information.
- Identify and interpret intercepts and zeros.
- Graph a linear equation.
- Identify equivalent equations.
- Solve an equation in one variable.
- Graph and interpret the line $y = x$.

Statistics: You will be able to

- Make and interpret a scatter plot.
- Make and interpret plots over time.
- Find a median fit line.
- Find a measurement for error in a line fit to paired data.
- Find a line that is the *best* fit for the data.
- Use the concept of residual to find prediction errors.
- Find and interpret an approximation for a correlation coefficient.

Unit 1

Linearity

EXPLORATORY LESSON

How Do Events Change Over Time?

How does the temperature in your state change from month to month?

From year to year?

Can you predict any patterns?

How does the cost of buying groceries change over time?

How does your height change from year to year?

Analyzing trends in events that occur over time is very important for people in many occupations in government, business, and industry, and for people who study climate, the environment, or agriculture. In this unit, you will explore ways different events change over time and develop mathematics necessary to describe and analyze the changes.

OBJECTIVE

Experiment with an outcome that will change over time.

EXPLORE

Practice Makes Perfect

Suppose you shoot a basketball for 4 minutes and record how many baskets you make in the first 30-second interval, in the second 30-second interval, and so forth, throughout the 4 minutes. Do you think that the number of baskets you make in each 30-second time interval will change over time?

Data Collection and Analysis

1. One hypothesis might be that as you continue to shoot baskets, you will improve your score with practice. Another

might be that as you continue to shoot, your score will become worse because you will grow tired.

a. Which of these two hypotheses do you think is more reasonable? Explain.

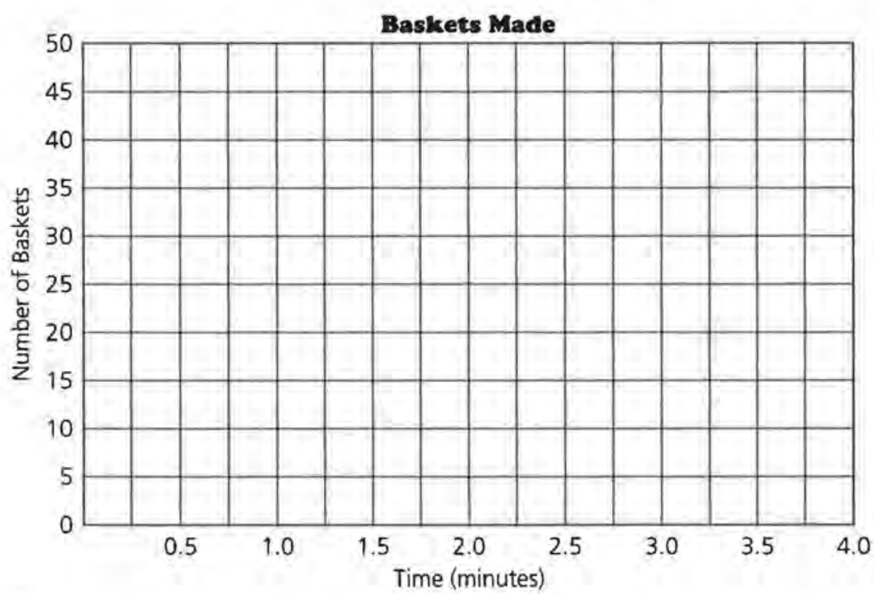
b. What is another possible hypothesis regarding your scores?

2. A graph showing the time intervals and the number of baskets can help you see the relationship between the length of time and the number of baskets you make.

a. What will the graph look like if the first hypothesis is true? If the second is true?

b. What will the graph look like if there is no relationship between the length of time you shoot and the number of baskets you make?

3. Conduct the experiment, and plot the results on a grid like the one below. Plot the time in minutes along the horizontal, or x-axis (0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0), and the number of baskets made along the vertical, or y-axis.



a. What conclusion can you make about the relationship between time and the number of baskets you made?

b. Do you think everyone in the class will have the same conclusion? Why or why not?

4. Collect and analyze the results gathered by several other students.

   a. How are the results alike? How are they different? How does your conclusion compare with those of your classmates?

   b. Determine a way to make a graph of the combined scores of the students in your group. Describe how you decided to use this method. How does each individual's score contribute to the overall results shown in the graph of the combined scores? What overall conclusion about time and scores can you make?

LESSON 1

Rates of Change

Are prices really going up?

Do all prices rise by the same amount?

If your salary rises at the same rate as the price of a car, can you buy the same model again next year?

OBJECTIVES

Find and interpret slope as a rate of change. Find the rate of change from data.

Often measurements change over time. For example, each year young children grow taller; the population increases or decreases; some salaries increase, and some salaries decrease; some prices rise, and some prices fall. It is important to understand change, not only to see what happened in the past but also to look for possible trends and patterns in the future. How much should you expect to pay for a car? How much can you afford to buy with a given salary? The problems in this unit will use mathematics to answer these questions.

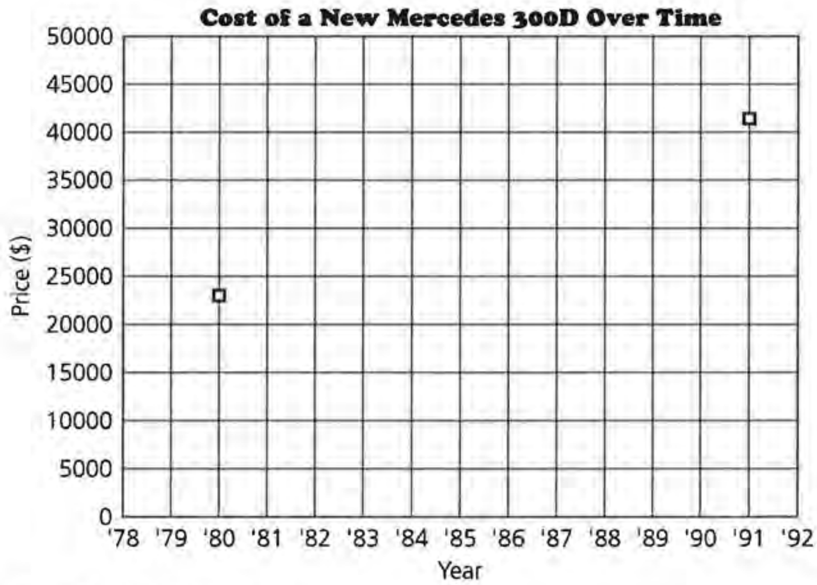
INVESTIGATE

New Car Prices and Requirements

In 1980, a brand-new Mercedes with a 300-diesel engine cost \$23,000. The price went up each year, so that in 1991, the price for the same car was \$41,350. How did the price change over time? Is there a way to describe this change per year?

Discussion and Practice

1. The prices of the Mercedes in 1980 and 1991 are plotted on the next page.
   a. What is the change in price from 1980 to 1991?
   b. If the price changes by the same amount each year, about how much is the change in price per year?
   c. Using your estimate, what was the price of the Mercedes in 1981?



Source: *Kelly Blue Book,* 1994

2. Assume that the change in price per year is represented by the constant amount found in the table below.

Year	Cost of Mercedes	Year	Cost of Mercedes	Year	Cost of Mercedes
1980	$23,000	1986	______	1991	______
1981	24,668	1987	______	1992	______
1982	26,336	1988	______	1993	______
1983	______	1989	______	1994	______
1984	______	1990	______	1995	______
1985	______				

a. Use *Activity Sheet 1* to complete the table and plot the points (*year, cost*) on the grid.

b. Describe the pattern of the points you see in the graph.

3. In 1985, a new Chevrolet Corvette sold for approximately $26,901; whereas in 1981, a new Corvette sold for $16,141.

a. Plot the two ordered pairs for the year and cost of the Corvettes. Using *only* your graph, estimate the change in price per year.

b. How does your estimate for the change in price per year of the Corvette compare with that of the Mercedes?

c. In 1991, a Corvette cost about $33,410. Plot this new point on your graph. Is the change in price the same, or *constant,* from 1981 to 1991? Explain how you found your answer.

Another way to describe the change in price per year is to calculate the rate of change in the price. Look at the following table of car prices.

New Car Prices

Year	Honda Civic	Chevrolet Camaro	Toyota Celica	BMW	Chevrolet Corvette	Mercury Cougar	Ford Mustang
1975	$2,799	$4,739	$3,694	$10,605	$9,424	$6,121	$4,906
1991	$9,405	$13,454	$14,658	$35,600	$33,410	$16,114	$11,873

Source: *Kelley Blue Book, 1994*

This is how to find the rate of change in price for a BMW.

$$\frac{35{,}600 - 10{,}605}{1991 - 1975} = \frac{24995}{16}$$

$$= 1562$$

The change in price for a BMW is about $1562 per year.

4. Use the table above.

a. Plot the points (*year, price*) for the Ford Mustang. Draw a line through the points, and estimate the rate of change in price per year.

b. On the same grid, plot the prices for the Toyota Celica. Draw a second line, and estimate the rate of change in price per year for the Toyota.

c. How do the two lines compare? How is the difference in the rates of change in prices reflected in the two graphs?

d. Look at *all* the data in the table. Do *all* of the cars have approximately the same rate of change in price over the time shown? Explain your conclusions.

5. Consider a set of data with a zero rate of change.

a. What do you think a *zero rate of change* means?

b. Sketch the graph of the cost of a new car for which the change in price over time is zero. Describe your graph.

6. According to the Recording Industry Association of America, there were approximately 53 million CDs (compact discs) sold in 1986. Each year thereafter, 58.5 million more CDs were sold than in the year before.

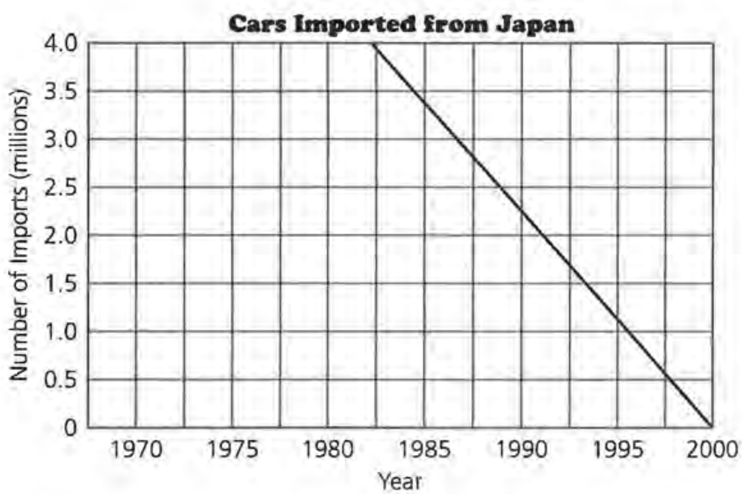
a. On grid paper, plot the data representing the number of CDs sold over time, and draw a line to summarize the relationship. What assumption did you make in order to draw your line?

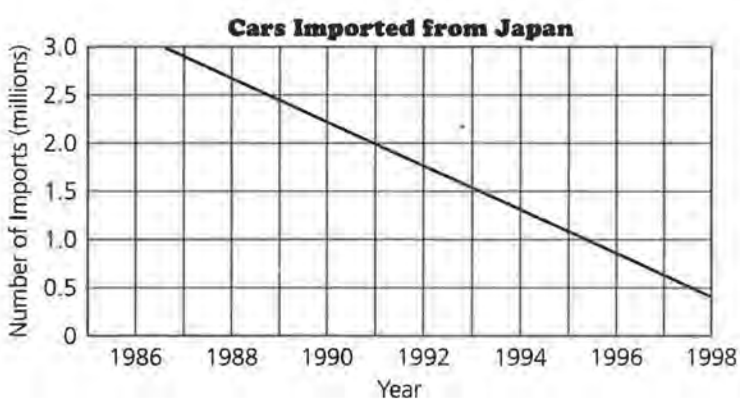
b. Using your graph, predict the number of CDs sold in 1993. Explain how you made your prediction.

c. Suppose the number of CDs sold per year increased by more than 58.5 million. What would your new line look like?

The *scale* used for a graph affects how you visualize the magnitude of a rate of change.

7. One newspaper reported that in 1988, 2.7 million cars were imported from Japan to the United States; whereas in 1992, 1.8 million cars were imported from Japan. A second newspaper reported that in 1988, 2.7 million cars were imported from Japan to the United States, and in 1990, 2.25 million cars were imported from Japan. Each newspaper article included a graph of a straight line representing the number of cars imported from Japan over the years, as shown here.





a. What is the rate of change in the number of cars imported from 1988 to 1992 as quoted in each newspaper? Explain the difference between the two graphs shown.

b. What can you conclude about the number of cars imported from Japan to the United States?

c. Plot the points (*year, number of imports*) on the *same* set of axes for both sets of data from the news articles. What conclusion can you now make? Do the articles contradict each other? Why or why not?

d. "To compare two rates of change, you either have to use mathematics or be very careful about your graph." What does this statement mean?

8. In 1991, the cost of a CD component with remote control was \$139. In 1989, the same component cost \$159.

a. On graph paper, plot the points representing the cost of the CD component each year.

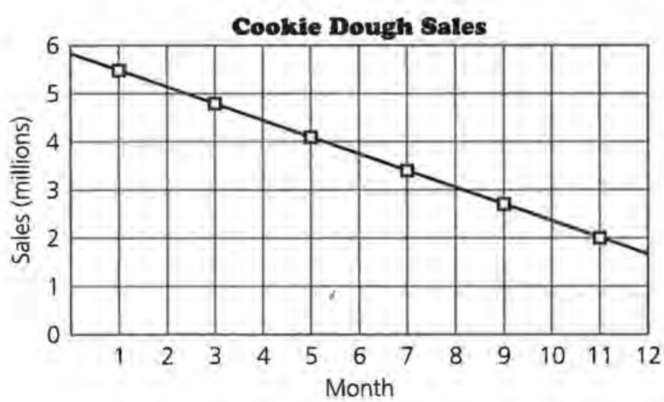
b. What is the change in price per year? What does it tell you?

c. What might explain the fact that the price of the CD player has decreased?

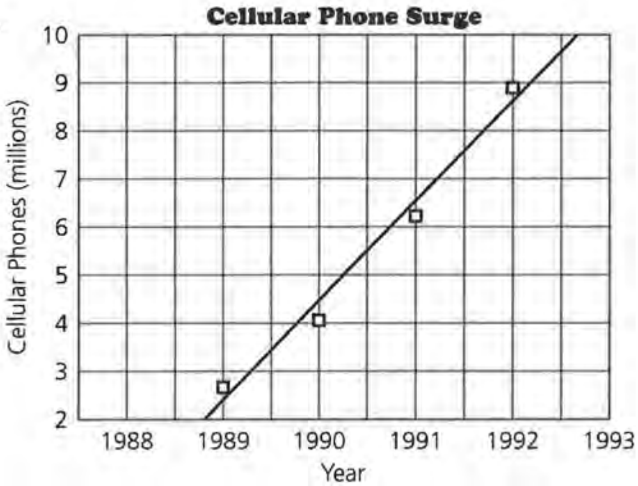
d. Name at least two other items whose prices have decreased over the past several years.

9. Describe, in words, the rate of change shown in each of the next two graphs.

a.

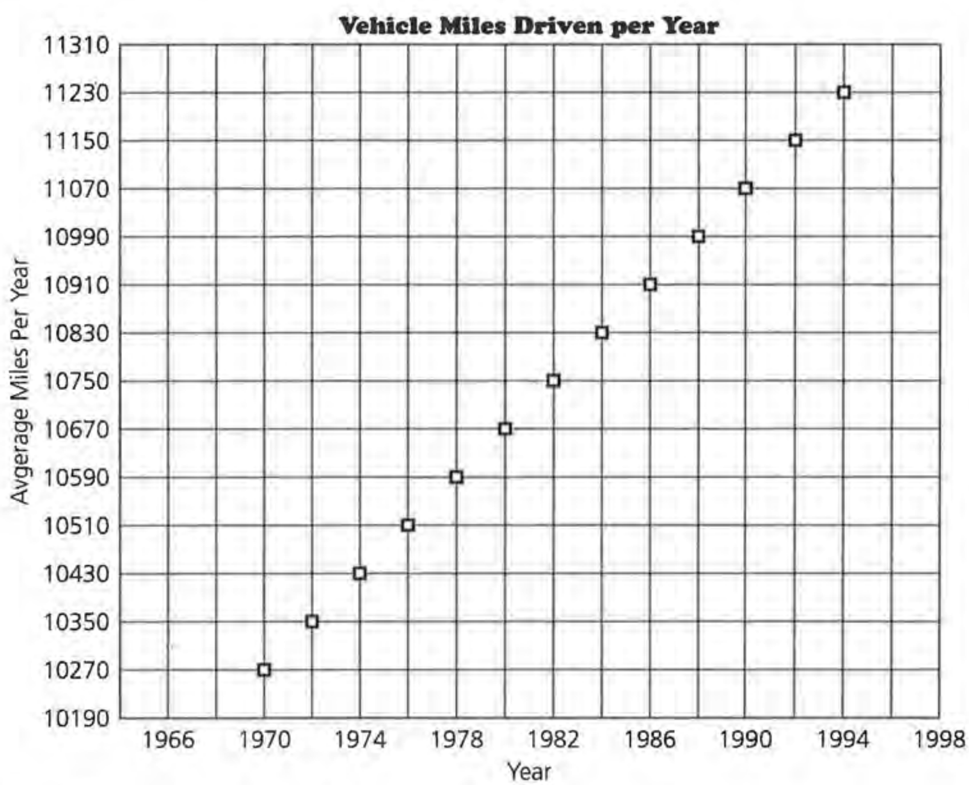


b.



Source: Cellular Telecommunications Industry Association, 1992

10. According to the 1993 Council on Environmental Quality, people drove a yearly average of 10,270 vehicle miles in 1970. The data for the estimated average number of vehicle miles per year, driven through 1994 are graphed below.



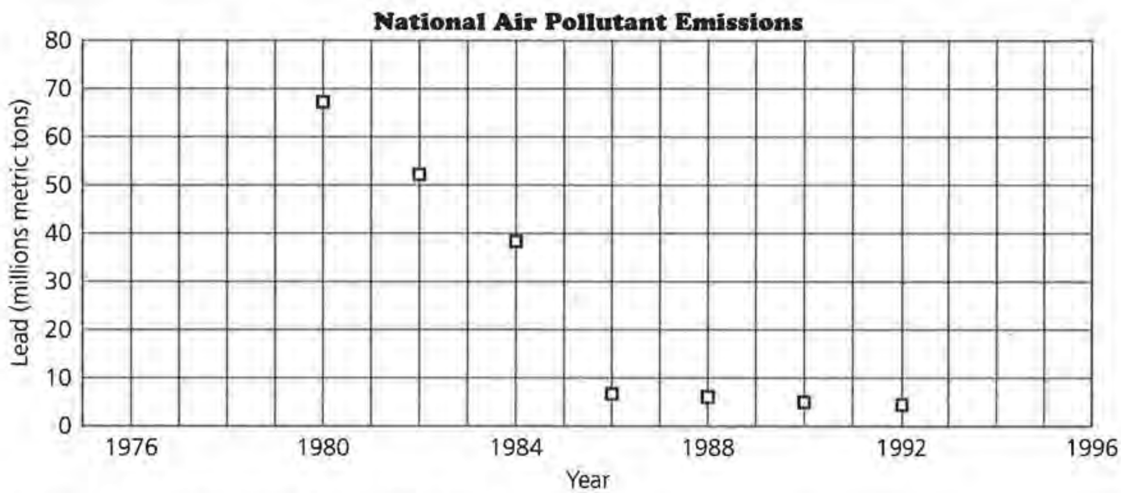
Source: data from 1993 Council on Environmental Quality

a. Find the average number of vehicle miles driven in 1974. What was the rate of change from 1970 to 1974?

b. What was the rate of change in the average number of vehicle miles driven from 1988 to 1994? How did you find your answer?

c. According to the report of the council, the average number of vehicle miles driven in 1993 was 11,090 miles. How does this value compare with the data shown in the graph? How did you find your estimate?

11. The amount of lead in the air from 1970 to 1992, based on information from the United States Bureau of the Census, is shown on the following graph. Recall that a graph of ordered pairs like this one is called a *scatter plot.*



Source: data from United States Bureau of the Census

a. Estimate the amount of lead in the air in 1980. What is the approximate rate of change from 1980 to 1984?

b. What was the rate of change in the amount of lead in the air from 1984 to 1986? How did you find the rate?

c. According to one environmentalist, the rate of change in the amount of lead in the air has *not* been constant, and in recent years it has been decreasing. Do you agree or disagree with that statement? Explain your reasoning.

d. For which scatter plot—the one for the amount of lead in the air or the one for the average number of miles driven per year—does it seem more reasonable to summarize the relationship with a straight line? How does the rate of change help you make your conclusion?

SUMMARY

When the rate of change is *constant* for equal time intervals, the graph of the relationship is a straight line. The *rate of change* is sometimes called the *slope* of the line. You can find the slope by finding the ratio of the change in the y-values of two representative data points and the corresponding change in the x-values of the same data points.

If (x_1, y_1) and (x_2, y_2) are two data points on a line, then the *slope* of the line is

$$\frac{\text{Change in } y}{\text{Change in } x} = \frac{\Delta y}{\Delta x} = \frac{y_2 - y_1}{x_2 - x_1}$$

(Note: Δ is a symbol used in mathematics and science to represent a change. For example, Δy means a change in the y-values.)

- The rate of change can be determined by any two points on the line.
- The rate of change can be positive, negative, or zero.
- If the rate of change is negative, the y-values decrease as the x-values increase. If the rate of change is positive, the y-values increase as the x-values increase; if the rate of change is zero, the y-values remain the same as the x-values increase.
- You can find the rate of change by reading a graph, by looking at the data in a table, or by finding the slope.
- For a straight line, the rate of change will be the same, or *constant,* between any pair of points on that line.

Practice and Applications

12. Jose finds the slope by using two data points as follows:

$$\frac{139 - 159}{1991 - 1989} = \frac{^-20}{2}$$

a. What ordered pairs (x, y) did he use?

b. Sue finds the slope by reversing the order of the same data points. Is the slope the same?

$$\frac{159 - 139}{1989 - 1991}$$

13. Lou's family purchased a used car that Lou was allowed to drive only to work. The car originally had 42,350 miles on it, and no one else in the family used it during the school year.

a. Lou worked at a car wash 10 miles from home. Plot on graph paper the point that represents the number of miles on the odometer after the first day Lou used the car.

b. Then plot the point that represents the mileage after Lou has used the car for 4 days. What is the rate of change in miles per day?

c. Continue to plot points that represent the mileage after any given number of days. Describe the pattern you see.

d. Calculate the rate of change in miles per day, first using the number of miles after 5 days of work and then again after 10 days of work. What can you conclude?

14. Gasoline prices are not constant but change continually depending upon demand, weather, and availability of gasoline. The price of gasoline at a given gas station did not change from Monday through Friday for a given week.

a. On graph paper, plot the prices for each weekday if the price was $1.12 on Monday.

b. During the rest of a 2-week period, the prices were as follows:

$1.15 on Saturday; $1.15 on Sunday; $1.12 on Monday; $1.10 on Tuesday; $1.10 on Wednesday; $1.18 on Thursday; $1.18 on Friday; $1.18 on Saturday; and $1.18 on Sunday. Plot the remaining data points for the 2-week period.

c. What was the largest drop in price? When did it happen?

d. What might cause prices to vary during this time?

15. The average daily price for gasoline at suburban New York City gas stations during the fall of 1990 is shown by the following graph.

Source: data from *The Milwaukee Journal*, October 27, 1990

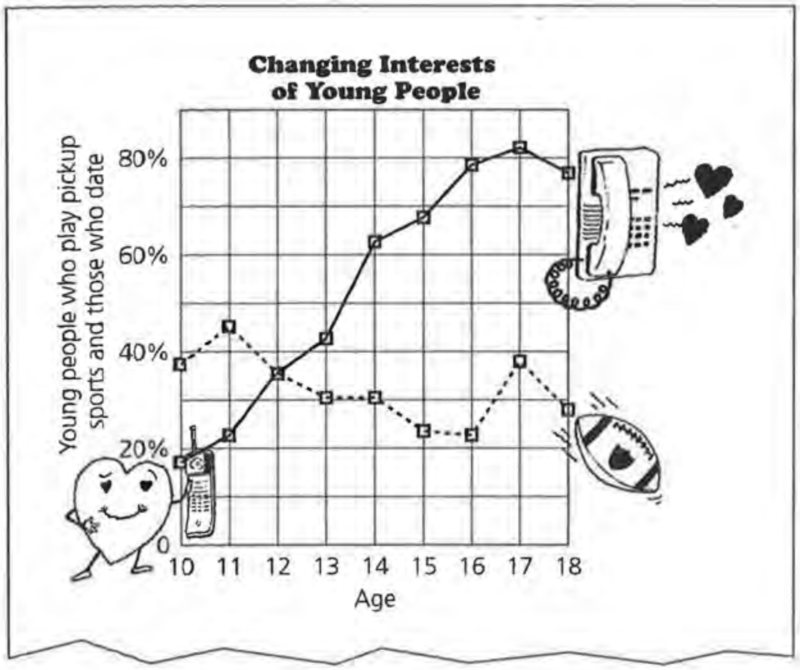
a. Describe how the graph was drawn.

b. What is your impression of the rate of change in the average price of gasoline per week and per day? Do you think the title of the graph is accurate? Why or why not?

16. For some data, the slope or rate of change is constant; for other relationships, the slope varies. How will graphs with these two types of slopes look? Draw an example of each.

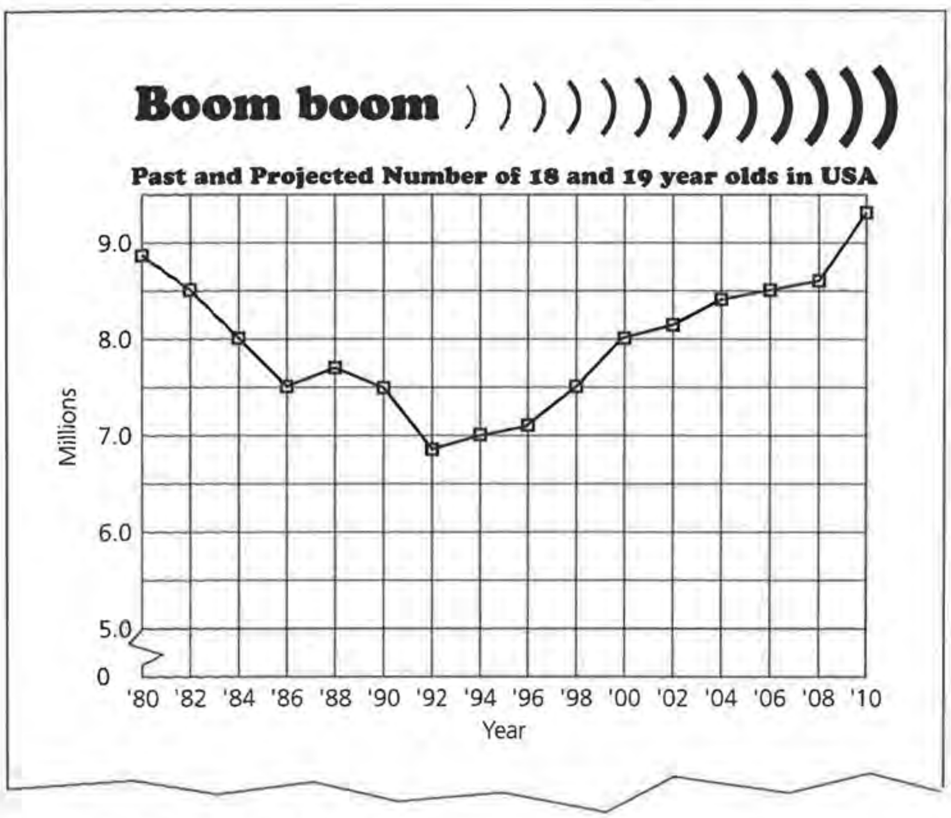
17. Describe the rate of change in each of the two graphs below.

a.



Source: *Youth Sports Institute*, Michigan State University

b.



Source: data from *USA TODAY*, April 19, 1995

18. The following graph and headline were used to describe the amount of business at fast-food restaurants.

a. Was the fast-food industry still growing in 1990? Explain your answer.

b. Suppose the total revenue for fast-food restaurants was \$200,000 in 1980. According to the graph, the growth in revenue in 1981 was 12.2%. Find the amount of revenue for 1981.

c. Some percentages of growth in revenue have been estimated from the graph and entered in the table on the next page. Estimate the remaining percentages to complete the table. Use the table on *Activity Sheet 1* to record the percentages.

Year	Revenue ($)	% Growth in Revenue	Growth in Revenue
1980	$200,000	———	———
1981		12.2%	
1982		10.0%	
1983		11.5%	
1984			
1985		8.0%	
1986		11.0%	
1987		9.8%	
1988			
1989		7.2%	
1990			

d. Was the amount of growth in revenue in 1990 (6.1%) half of that in 1981 (12.2%)? Explain why or why not.

e. Draw a graph of the *amount* of revenue. The graph in the article shows the *rate* of growth in revenue. How does your graph compare to the one shown?

19. Find the slope, or rate of change in price, for each pair of points plotted on the following two grids. What real-world situations could the points represent? Describe what each rate of change tells you.

a.

b.

20. Draw a graph that reflects the change in the number of teeth a person has over his or her lifetime. Explain why your graph looks the way it does.

21. Study the following tables.

Year	Number of Refugees in United States (millions)
1980	8
1981	10
1990	16
1992	18

Year	% of Working Married Women with Children Under Age 6
1965	23%
1970	30%
1975	37%
1990	59%

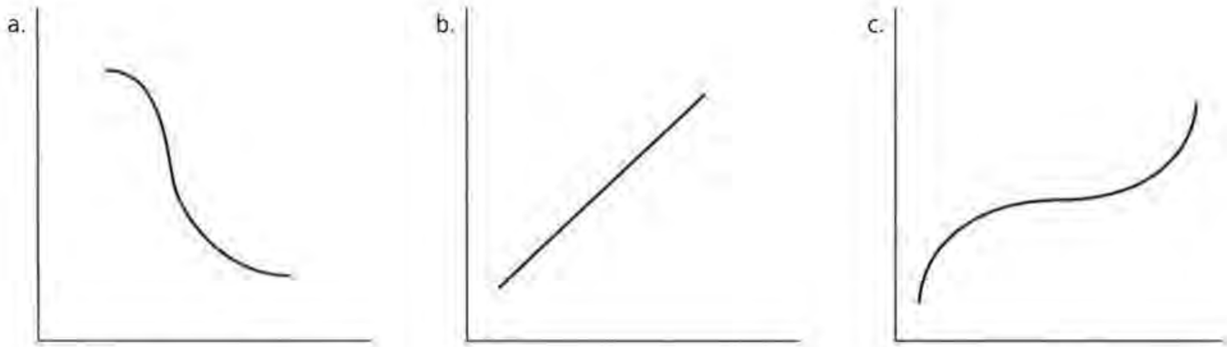
Year	Number of College Football Games Cable Cast
1987	56
1988	62
1989	98
1990	192

Sources: Worldwatch Institute, United Nations High Commission for Refugees; *USA Today,* September 2, 1993; and Federal Communications Commission Interim Report, 1992.

a. Does the rate of change appear to be constant for each table?

b. Which, if any, of the graphs for the data sets in the tables above do you think will be straight lines? How did you decide on your answer?

22. The following graphs show the sales of three types of shoes—tennis shoes, boots, and dress shoes—over the past 10 years. Match each graph with a shoe type. Explain your choices.



23. Describe each of the following:

a. a line with positive slope.

b. a line with negative slope.

c. a line with zero slope.

24. Ask an adult to approximate how much an item such as a candy bar or a loaf of bread cost at least 10 years ago, then find how much the same item costs today. On graph paper, plot the data, calculate the slope, and predict how much you think the item will cost in the year 2000. How reliable do you think your prediction is?

25. **Challenge** *Inflation* is the increase each year in the cost of goods and services. Suppose the rate of inflation has averaged approximately 3% per year. Calculate how much a 1991 Corvette would cost if its price had increased by the average rate of inflation each year since 1975, when it cost \$9424. Record the data in a table.

a. Make a scatter plot of these points.

b. How does the 1991 "rate of inflation" price compare with the actual price in Question 3 on pages 7–8? Why do you think the prices are different?

c. Is the rate of change in the price of an item due to inflation constant? How did you decide?

LESSON 2

Equations of Lines

How do you think the prices of new cars have actually changed over the years?

What will the graph look like if you include the prices from several years?

Do you think the rate of change has actually been constant?

In Lesson 1, you investigated lines using only two data points. In doing so, you had to assume that the line drawn through the two points represented all of the data. If you have more data points, you can study the scatter plot for any overall pattern and determine if most of the points lie near one straight line.

OBJECTIVE

Write the equation of a line drawn through a set of points that appear to be linear.

INVESTIGATE

Buying a Honda Civic

The list prices for several types of cars are shown in the table on the next page.

Discussion and Practice

1. Look at the prices for the Honda Civic.

   a. What pattern do you think a scatter plot of the prices for a new Honda Civic for the given years will show?

   b. The following scatter plot shows the prices of the Honda Civic. How does the pattern you see reflect your guess?

New Car List Prices

Year	Honda Civic	Chevrolet Camaro	Toyota Celica	Ford Mustang	Mercury Cougar*	BMW	Chevrolet Corvette
1971	$1,395	$3,790	$2,847	$3,783	$4,069	$5,845	$6,327
1973	2,150	3,829	3,159	3,723	4,045	8,230	7,007
1975	2,799	4,739	3,694	4,906	6,121	10,605	9,424
1977	2,849	5,423	5,252	4,814	6,225	14,840	11,508
1979	3,649	6,021	6,904	5,339	6,423	20,185	12,550
1981	5,199	8,142	7,974	7,581	8,762	24,605	16,141
1983	4,899	9,862	8,824	8,466	10,725	24,760	N/A
1985	6,479	10,273	9,549	8,441	11,825	25,360	26,901
1987	7,968	11,674	12,608	9,948	14,062	29,220	28,874
1989	9,140	13,199	13,528	11,145	15,903	37,325	32,445
1991	9,405	13,454	14,658	12,321	16,890	35,600	33,410
1993	8,730	16,385	15,983	12,847	17,833	38,355	36,230
1995	10,130	17,536	19,410	17,550	18,960	39,775	37,955

Source: *Kelley Blue Book*, 1996.
*Prices are for V8 or V6 depending on model made that year.

Source: *Kelly Blue Book*, 1996

2. Suppose you draw a line representing the pattern you see in the plotted data.

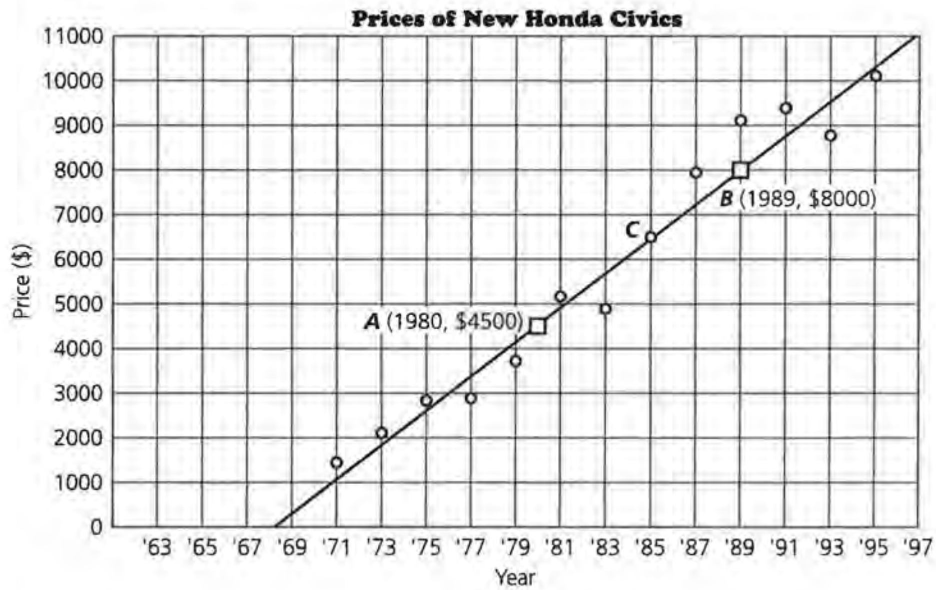
a. How many points are necessary to determine a line?

b. Bashom drew his line through the point for 1971 and the point for 1993. Do you think his technique was a good one? Explain why or why not.

c. Audrey drew her line so that it went through as many points as possible. What do you think about her method compared with Bashom's method?

When the points show a pattern similar to the one above, they are approximately *linear* over the time interval given. A line can be drawn close to the points summarizing the relationship of the ordered pairs of data.

3. The following scatter plot shows a line drawn to approximate the pattern in the prices of the Honda Civic over time. The ordered pairs (1980, \$4500) and (1989, \$8000) are labeled A and B, respectively.



a. Are any of the actual data points on the line? If so, which ordered pairs do the points represent?

b. Using the line, estimate the price of a Honda in 1989. What is the difference between the estimated price and the actual price?

c. Using *only* points that lie on the line, estimate the annual rate of change in the price of a Honda.

4. In 1980, a new Honda cost about \$4500. Two years later, the price of a new Honda was about \$778 more than it was in 1980.

a. Using the graph from Question 3, find the cost of the Honda in 1982. How does this value compare with an increase of \$778?

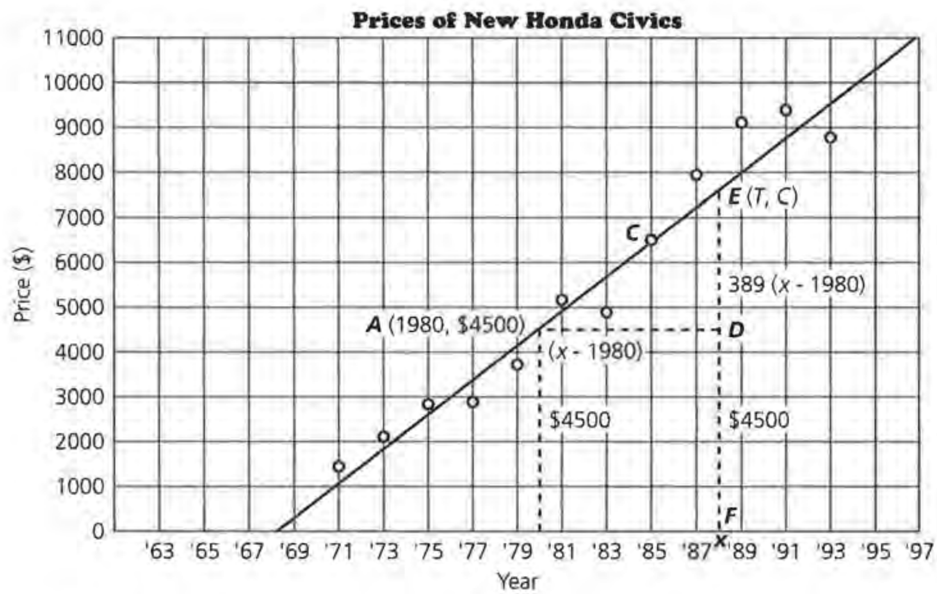
b. Complete the table at the top of the next page using *Activity Sheet 2*. Use your completed table to help you write a general rule or equation for the cost. Be sure to indicate what the variables in your equation represent.

Cost	Calculations	Dollar Amount
Cost in 1980 =	4500 + 389 (0)	4500
Cost in 1983 =	4500 + 389 (_______)	_______
Cost in 1979 =	4500 + 389 (_______)	_______
Cost in 1988 =	4500 + 389 (_______)	_______
Cost in General =	4500 + 389 (_______)	_______

c. Use your rule or equation to estimate the cost of a new Honda in 1994.

d. If you use the rule "The cost of a new Honda equals the initial cost plus the increase in cost," how can you find the increase in cost?

5. Use the following scatter plot to help you analyze the equation or rule you wrote to describe the general price of a Honda based on the price in 1980.



a. What are the coordinates of point *D*?

b. What does *DF* represent? (Recall that *DF* means the length of the line segment *DF*.) What does *ED* represent?

c. How does *EF* relate to the price of a Honda?

d. Use the graph to help you describe what the numbers represent in this equation:

Cost = \$4500 + \$389 (9) = \$8001

6. Sara wrote this equation to describe the price of a new Honda:

Cost = \$4500 + \$389(T – 1980)

a. In Sara's equation, what does T represent? Describe in words what the numbers \$4500, \$389, and 1980 represent.

b. Use Sara's equation to estimate the price of a new Honda in 1991.

c. How does the price from the line compare with the actual price shown as a data point on the scatter plot?

d. Use Sara's equation to make a table for the price of a Honda Civic in even-numbered years. Using the table and grid shown on *Activity Sheet 2*, describe a quick way to generate the data in the table, and then plot these points on the grid. Where do the points lie on the scatter plot?

Year	Cost in \$
1970	______
1972	______
⋮	______

7. When the ordered pair (1980, \$4500) is used as the initial value for the price of a Honda, the equation is C = \$4500 + \$389(T – 1980).

a. Rewrite the equation when the initial value is (1989, \$8000).

b. Use each equation to predict the prices of a Honda in 1995. How do the results compare?

8. Suppose the equation for the cost of a new Mercedes is

C = \$23,000 + \$1668(T – 1980).

a. What do the numbers in the equation represent?

b. According to the equation, approximately how much did a new Mercedes cost in 1991?

c. Using these ordered pairs, on graph paper draw a scatter plot of the prices of a new Mercedes over time. How do these prices of a Mercedes compare with the data in Question 2 from Lesson 1?

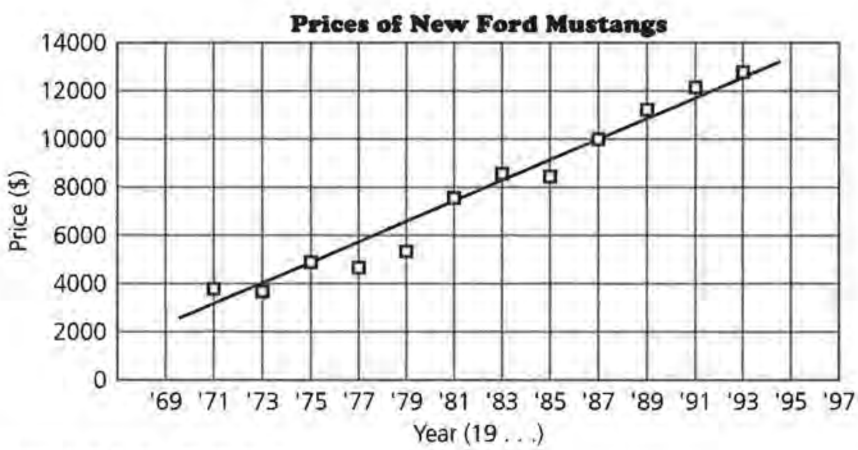
SUMMARY

Points that form a straight line can be represented by a ***linear equation.***

- To summarize linear patterns, you can draw a straight line approximating the pattern shown in the scatter plot of the data.
- The relationship shown by the line can be written as a rule or equation.
- To write the equation of the line determine an initial value (t, c) and a rate of change.
- An equation of the form $C = c + r\,(T - t)$ can be used to represent the relationship between cost and time in which T is any year in general, t is a specific year, C is the cost in general, and c is the cost in the specific year.
- In general, if the ordered pair (x_1, y_1) is the initial value, and m is the rate of change or slope, then the equation of the line is $y = y_1 + m(x - x_1)$.
- The equation of the line can be used to generate any point on the line. Every point on the graph of the line satisfies the equation.

Practice and Applications

9. The following scatter plot shows the actual prices of a new Ford Mustang for every other year from 1971 to 1993. Look carefully at the line drawn on the scatter plot.



Source: *Kelly Blue Book,* 1996

a. Find the point (1988, \$10,500). Find another point on the line, and calculate the rate of change in price for a new Mustang. What does this tell you about the price of a new Mustang over the years?

b. Use the ordered pairs to write an equation for the line.

c. Find a third point on the line. Substitute the ordered pair into your equation. Is your equation a true statement?

d. Use your equation to predict the price of a new Mustang in 1992. How close to the actual price do you think your prediction will be?

e. What price does your equation predict for the cost of a new Mustang in 1995? What is the difference between the prediction and the actual cost of a Mustang in 1995? What might explain the difference?

10. Have each group member choose one of the following from the table at the beginning of this lesson on page 22: Camaro, Celica, Cougar, BMW, or Corvette. Plot the points for the years and the prices since 1971, and draw a line to summarize the relationship if it seems appropriate. Compare the scatter plots of all of the cars and use them to answer the following questions. Be sure to choose the same scale for all of the graphs so you can compare the scatter plots easily.

a. Estimate the rate of change in the price of the car you selected, and explain what it means.

b. Compare your graph with the others in your group. For which car does a constant rate of change seem most appropriate? How does your graph differ from the others?

c. Do any cars have a rate of change in price that does not seem very constant? Explain how you decided on your answer.

d. Which car's cost has the greatest rate of change? How does its graph differ from the others?

e. Write an equation for each line using the year 1975 as the starting year. What observations can you make?

11. Sales for Outlook Graphics, a company that sells baseball cards, were worth \$7.7 million in 1986 and \$65 million in 1992. The change in price of baseball cards was relatively constant.

a. Write an equation to describe the sales for the company over the years.

b. What do your variables represent?

c. Use your equation to predict the sales of the baseball cards in 1995.

12. The cost of a share of stock in Outlook Graphics in September 1991 was \$11.75. At the end of January 1992, it was \$22.00.

a. The price of a share rose at about the same rate each month. Write an equation that describes the cost of the stock over the months. What do your variables represent?

b. Describe the slope, or rate of change, in terms of price and time.

c. Predict the cost of a share of stock in 1994. What assumptions did you make?

13. The following table gives the number of firearm homicide victims in the United States, according to the Federal Bureau of Investigation.

Firearm Homicide Victims

Year	1988	1989	1990	1991	1992
Number of Victims	10,900	12,000	13,000	14,200	15,400

Source: *Statistical Abstract of the United States, 1995*

a. Plot the points and draw a line that you think will summarize the relationship between time and the number of homicide victims.

b. Write an equation describing this relationship.

c. How many firearm homicide victims would you expect in 1993?

LESSON 3

Equivalent Forms of Equations

In 1992, approximately 92.1 million households had television sets. During that year, major league baseball club owners decided to find out how many of these people were watching televised major league games. Were TV baseball ratings declining?

What kind of audience is predicted for the future, and what does this mean to baseball club owners and to the television networks?

Baseball club owners wanted to predict the trend in the number of people watching baseball games on TV. The number of viewers affects the ratings; the ratings affect how much television networks are willing to spend to purchase the rights to televise the baseball games.

OBJECTIVE

Determine if two equations are equivalent by using graphs, ordered pairs, and equivalent forms.

INVESTIGATE

Watching Baseball on TV

The table that follows on the next page displays information from the Federal Communications Commission interim report. The report shows the average TV ratings of nationally televised regular season major league baseball games since 1980. Suppose the data are linear; a line has been drawn to summarize the pattern. If two of the owners used different sets of points on that line to find its equation, would the equation be the same? Stated another way, does every line have only one equation?

Regular Season TV Ratings for Major League Baseball

Year	Ratings in Points
1980	8.0
1981	6.5
1982	8.1
1983	8.0
1984	7.0
1985	6.5
1986	5.5
1987	6.6
1988	6.0
1989	5.0
1990	4.6
1991	4.0
1992	3.4

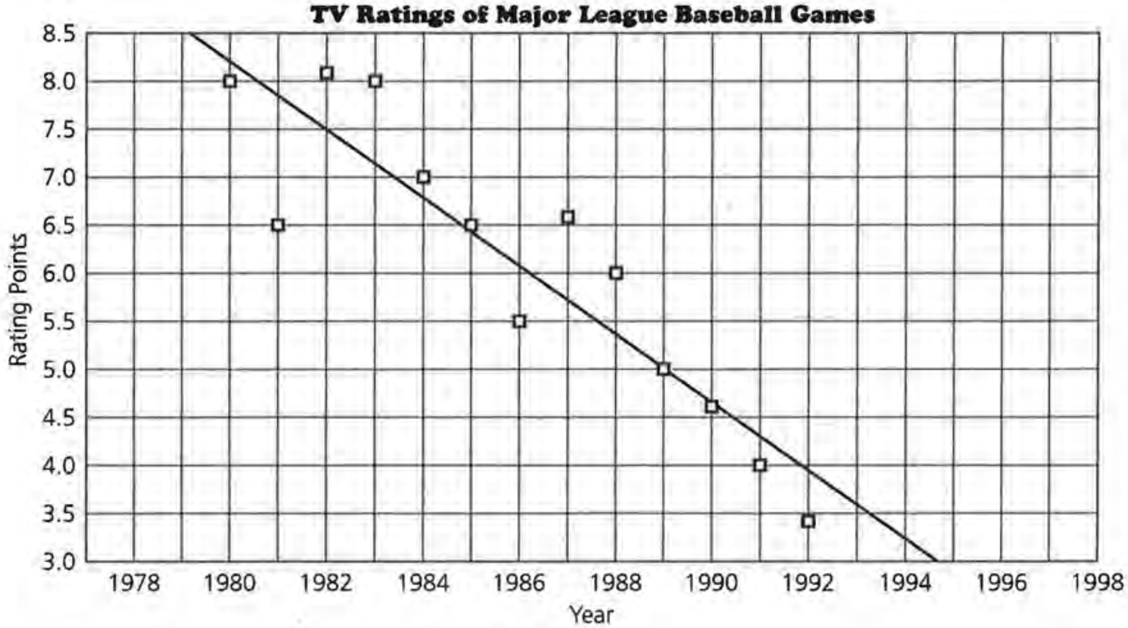
Source: *Federal Communications Commission Interim Report,* July 1, 1993

Discussion and Practice

1. Why do you think the networks and baseball club owners would be concerned about the number of TV viewers?
2. One rating point means approximately 1% of the TV households in the United States. How many households could have been watching major league baseball in 1992?

Suppose you have two equations, and you would like to know whether or not they represent the same line. Remember that you can find sets of ordered pairs for any linear equation that make the equation true. The graph of every linear equation is a straight line. If two equations represent the same graph, they are called ***equivalent equations***.

3. A scatter plot of baseball TV ratings with a line used to summarize the data, is shown here.



Source: *Federal Communications Commission Report*, July 1, 1993

a. Find two points on the line, determine the rate of change, and write an equation for the line.

b. Compare your equation to the ones found by your classmates. Describe how you could determine whether or not all of the equations represent the line shown in the scatter plot.

4. Consider these two equations: $c = 20 + 2(x - 4)$, and $c = 28 + 2(x - 8)$.

a. For each equation, find the rate of change or slope and a point (x, c) on the line determined by that equation.

b. Graph each equation. How do graphs help you determine whether or not two equations are equivalent?

5. Let $x = 10$. Find c for the equations $c = 20 + 2(x - 4)$ and $c = 28 + 2(x - 8)$.

a. Are the two equations equivalent? Justify your response.

b. Generate tables of ordered pairs (x, c) that satisfy both equations, beginning with $x = 0$. Compare the two tables. What conclusion can you make?

c. How many identical ordered pairs do you need before you can determine whether or not the equations are equivalent?

6. Consider $c = 18 + 2(x - 2)$ and $c = 12 + 4(x - 3)$.

a. Let $x = 7$ for both equations, and find c. Are the two equations equivalent? Justify your response.

b. Make tables of ordered pairs and graph these two equations to determine whether or not the equations are equivalent.

c. How do ordered pairs help determine whether or not two equations are equivalent?

To determine if two equations represent the same line and the same set of ordered pairs, you can rewrite each equation in a different form using algebraic properties. Distribute and combine like terms in each equation, and then compare the forms. If the forms for the equations are identical, the equations represent the same line. Rewriting the equations is an efficient way to determine whether or not two equations are equivalent.

Are $c = 20 + 2(x + {}^{-}4)$ and $c = 28 + 2(x + {}^{-}8)$ equivalent equations?

$c = 20 + 2(x + {}^{-}4)$		$c = 28 + 2(x + {}^{-}8)$
$c = 20 + 2x + {}^{-}8$	Distribute.	$c = 28 + 2x + {}^{-}16$
$c = 12 + 2x$	Combine like terms.	$c = 12 + 2x$

7. Do the two equations above represent the same line? Why or why not?

8. Using the **Distributive Property**, rewrite each of the following so that the expression is in the form $a + bx$ for some numbers a and b. Combine like terms where possible.

a. $2(x + 5)$

b. $3(x - 4)$

c. $10 + 5(x - 9)$

d. ${}^{-}4 + 2(x + 3)$

9. Consider these two equations: $r = 26 + {}^{-}12(x - 2)$ and $r = 70 + {}^{-}12(x - 10)$.

a. Can you find an equivalent form for each equation? Do the two equations represent the same line?

b. Why do you think using equivalent forms is a useful way to determine whether or not two equations are equivalent?

10. Use the equation of the line you found for the baseball ratings in Question 3 on page 31. Rewrite your equation in the form $y = a + bx$, in which a and b are numbers, x represents time, and y represents rating points. Compare your results with the results of your classmates.

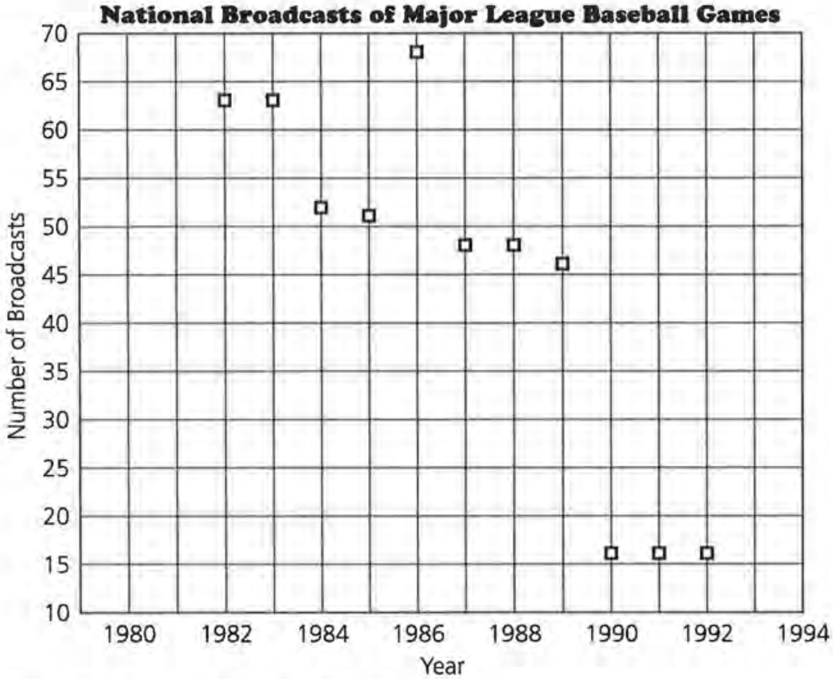
11. As the baseball club owners and TV networks analyzed the situation in 1993, they looked at other information about baseball and television viewing. In particular, they looked at the numbers of local telecasts and national telecasts of regular season baseball games as shown in the following table.

Televised Baseball Games

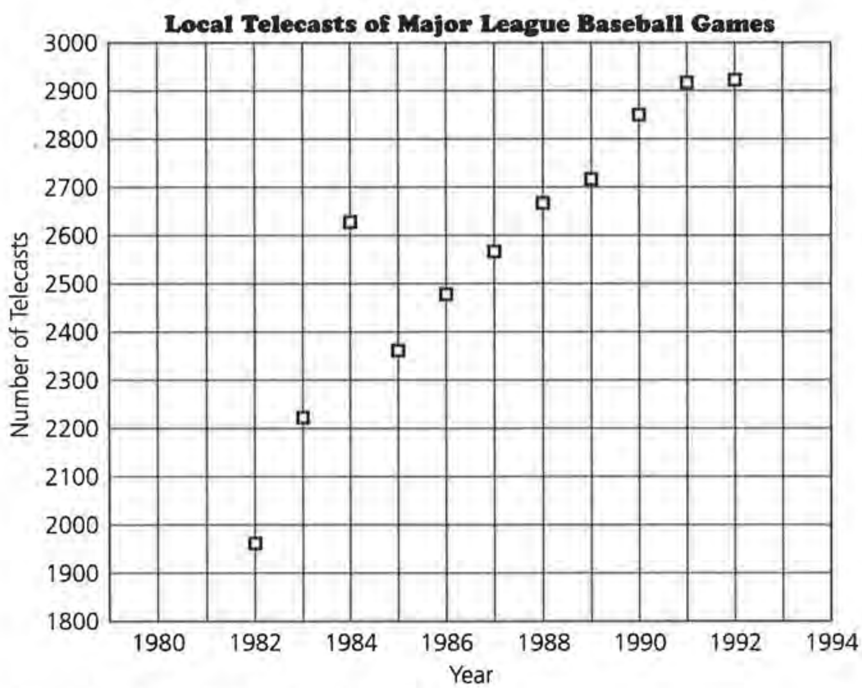
Year	National Telecasts	Local Telecasts
1982	63	1954
1983	63	2217
1984	52	2623
1985	51	2356
1986	68	2473
1987	48	2568
1988	48	2661
1989	46	2714
1990	16	2849
1991	16	2917
1992	16	2922

Source: data from *USA Today,* July 23, 1993

a. Look at the graphs of the data in the following two scatter plots. What are some significant differences between the scatter plots of the two sets of data?



Source: data from *USA Today,* July 23, 1993

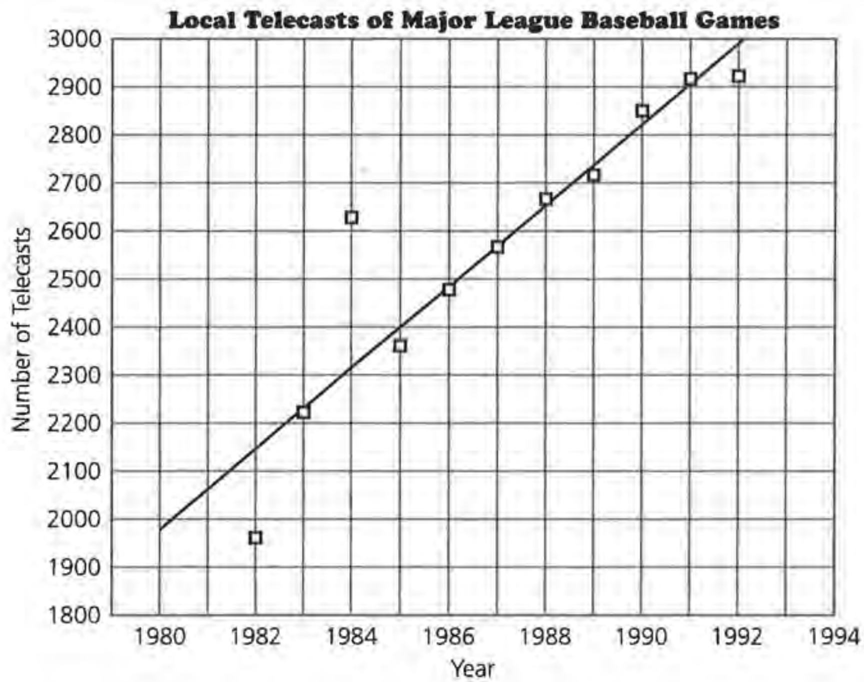


Source: data from *USA Today,* July 23, 1993

b. Explain whether or not it makes sense to draw a line to summarize either relationship.

c. What connections might there be between these data and the data on TV viewing ratings at the beginning of this lesson on page 30?

12. A line summarizing the data for the number of local telecasts is shown on the following scatter plot.



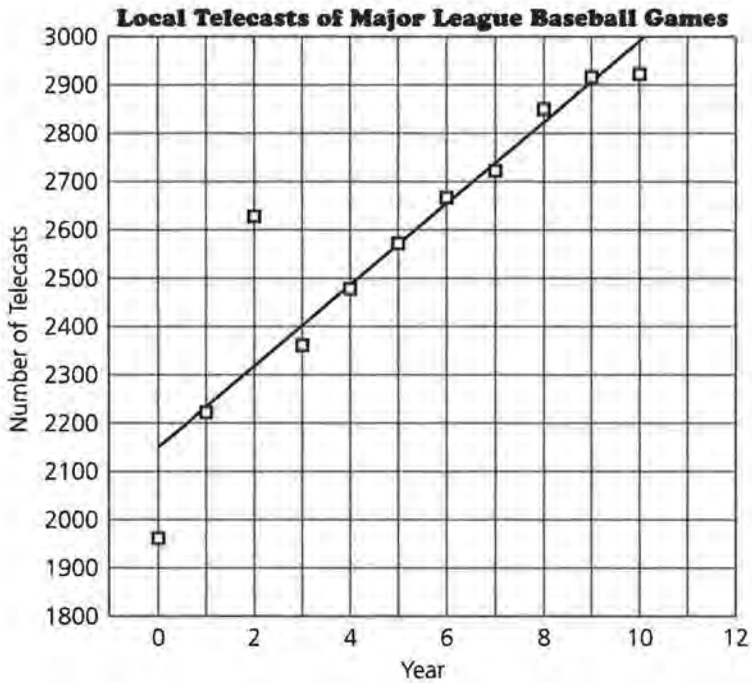
a. Using the line, how many local telecasts were there in 1984? How would you describe the data point for that year?

b. Write an equation for the line. What do your variables represent? Describe the rate of change in the number of local telecasts over time.

c. Compare your equation with the equation of a classmate. Are the two equations equivalent? How can you decide?

d. Write your equation in the form $y = a + bx$. Use your equation to predict the number of local telecasts in 1994. Compare your prediction with those of your classmates.

13. The following is a different scatter plot of the number of local telecasts.



a. How is this scatter plot like the scatter plot in Question 12 on page 35? How are the two scatter plots different?

b. Write the equation of the line shown in this scatter plot. What do your variables represent?

c. Write your equation in the form $y = a + bx$.

d. How is this equation like the one you wrote before? How are they different?

14. You have looked at several different data sets about television and major league baseball. What advice would you have given to the owners and networks in 1993, based on the analysis you have done in this lesson?

SUMMARY

Two equations are **equivalent** if they

- Have exactly the same graph.
- Generate the same set of ordered pairs given equal slopes.
- Can be rewritten in exactly the same form.

If two equations have the same slope and share a common point, they are equivalent. An equation of the form

$y = a + m(x - c)$ for some numbers a, m, and c and variables (x, y) can be rewritten as $y = d + mx$, in which $d = a - mc$, by distributing and combining like terms. In both forms, m represents the rate of change or slope.

Practice and Applications

15. Recall the conditions under which two or more equations are equivalent.

a. Which of the following equations are equivalent?

Bob: $y = 18 + 0.5(x - 2)$ Suzie: $y = 20 + 0.5(x - 6)$

Tom: $y = 22 + 0.5(x - 10)$ Lee: $y = 16 + 0.5(x - 1)$

Lashanda: $y = 19 + 5(x - 3)$ Pat: $y = 34 + 2x$

Sol: $y = 2x - 34$

b. Which of these equations determine the same straight line? How did you decide?

16. The annual fees paid by TV networks to televise major league baseball were \$46 million in 1980 and \$365 million in 1990. Assume the rate of change has been relatively constant over the years.

a. Write an equation that describes the relationship between the annual fees and time. What is the slope of your equation, and what does it tell you about the fees?

b. Use your equation to estimate the annual fees in 1993.

c. In 1993, the cost of a 30-second commercial during the telecast of a regular National Baseball League game was \$60,000. How many 30-second commercials would the network show to cover the cost of their fees?

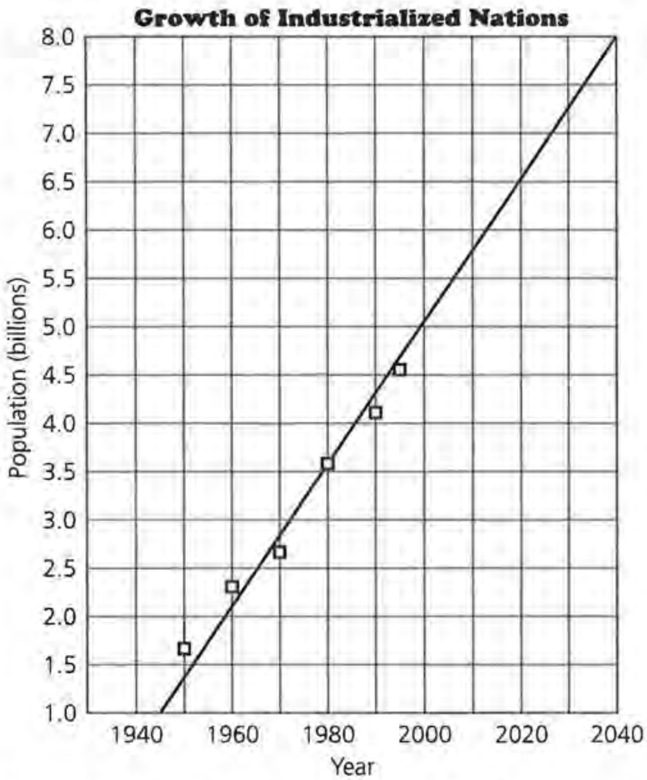
17. Kuong says the line determined by $y = 10 + 3(x - 2)$ contains the point (10, 2) and has a slope of 3. Is he correct? Why or why not?

18. If two lines have the same slope, what do you know about the lines?

19. If the two equations $y = a + bx$ and $y = c + dx$ represent the same line, what can you say about the numbers a, b, c, and d? Give an example to justify your answer.

20. The growth and projected growth of the population in the industrialized nations of the world, according to the United

Nations Population Fund, is shown on the following scatter plot. A line has been drawn to summarize the data.



Source: data from United Nations; Population Reference Bureau

a. Name some countries that you think would be included as industrialized nations.

b. Write an equation you could use to make projections about the population. Use this equation to estimate the population in industrialized nations in 1990. The actual population according to the United Nations was 4.086 billion. What is the difference between your estimate and the actual population?

c. Predict the population for the year 2020, using the line shown. The projected population according to the United Nations is 8.2 billion. What might explain the difference between the two projections?

d. In general, how reliable do you think these projections will be? Give a reason for your answer.

21. Not all scatter plots of data over time will show a linear pattern. Look through newspapers or magazines to find at least two examples in which data graphed over time are not linear. Provide descriptions along with your examples.

ASSESSMENT

Price Changes and Median Earnings

OBJECTIVE

Apply the concepts of slope and lines.

Project: Price Changes

1. Choose five things you have purchased several times over the past four years. Make a graph for each item showing the year and prices.
   a. Estimate the rate of change in price for each item.
   b. Compare the changes in price over time.
2. Think of several situations in which data change, but the rate of change is not constant. Describe and sketch a graph for each example.

Assessment: Median Earnings for Men and Women

The median earnings of year-round, full-time workers age 25 years and older for men and women are given in the following table.

Median Earnings

Year	Men	Women
1970	$9,521	$5,616
1971	10,038	5,872
1972	11,148	6,331
1973	12,088	6,791
1974	12,786	7,370
1975	13,821	8,117
1976	14,732	8,728
1977	15,726	9,257
1978	16,882	10,121
1979	18,711	11,071
1980	20,297	12,156
1981	$21,689	$13,259
1982	22,857	14,477
1983	23,891	15,292
1984	25,497	16,169
1985	26,365	17,124
1986	27,335	17,675
1987	28,313	18,531
1988	28,180	18,509
1989	29,556	19,752
1990	29,987	20,556
1991	30,874	21,272
1992	31,408	22,141

Source: *U. S. Census Bureau,* 1995.

1. Study the table above. Compare the earnings of men with those of women over time.

2. Plot the median earnings of men since 1970. Along the horizontal axis graph the year, and along the vertical axis graph men's earnings. Draw a line to represent the relationship.

a. Write an equation that relates the year and the median earnings of men.

b. What does the rate of change tell you about men's earnings?

c. Predict how much a man would have earned in 1993, according to your line. Circle this point on your graph. How accurate do you think your prediction is?

d. The median earnings in 1993 for a male was $31,600. What is the difference between this value and your prediction?

3. The median earnings of women are also given in the table. Plot the earnings data for women on the same grid you made for men's earnings. Draw a line that represents this relationship.

a. How do the two lines compare? How do the two rates of change compare?

b. Write an equation that can be used to predict, in general, the median earnings for women. Predict the median earnings for women in 1993.

c. The actual median earnings for women in 1993 was $22,422. How close to this value is your prediction?

4. A local newspaper article reported that women's salaries are growing at a faster rate than men's salaries. Use the graph and your work to refute or defend the statement.

5. **Bonus** Describe another way to represent both men's and women's earnings, using a scatter plot. Explain how this representation can be used to compare men's and women's earnings over time.

LESSON 4 (OPTIONAL)

Buying Power

Has the change in what you can buy with the money you earn been constant?

"Buying power" refers to the percentage of income needed to purchase an item. What would a graph of your buying power look like?

Unless earnings and the cost of living change at exactly the same rate, buying power will vary. By analyzing data, you can determine how much buying power has increased or decreased over time.

OBJECTIVES

Solve problems by applying the concepts of *percentages, slope,* and *linearity.* Analyze data that do not have a constant rate of change.

INVESTIGATE

Earnings, New Cars, and Groceries

Using the median earnings data for men, how much of his salary did a man spend to buy a new Honda Civic in 1981? How much did he spend in 1991 to make the same purchase? What has happened to his "buying power" in the past 11 years? How are these two sets of data related: median earnings of men and women, and new car prices over time?

Discussion and Practice

1. Use the car prices data at the beginning of Lesson 2 on page 22 and the earnings data at the beginning of the Assessment on page 39 to answer the questions.
   a. In 1981, the median earnings for a man was $21,689, and a new Honda Civic cost $5199. What percentage of a man's income would be spent to buy a Honda?
   b. Find the percentage of a man's income spent to buy a new Honda in 1991. How do the two percentages for 1981 and 1991 compare?

c. With respect to buying a new Honda, how has men's buying power changed?

2. The table shows the average weekly cost of groceries for a family of four in the Milwaukee, Wisconsin, area.

Grocery Cost and Percentage of Men's Median Income

Year	Weekly Cost of Groceries	% Income for Year	Year	Weekly Cost of Groceries	% Income for Year
1981	$59.00		1988	$47.00	
1982	61.00		1989	51.50	
1983	56.00		1990	49.50	
1984	45.00		1991	51.00	
1985	45.00		1992	55.00	
1986	44.50		1993	50.00	
1987	47.00				

Source: *The Milwaukee Journal Consumer Analysis, 1981-1993.*

a. Using *Activity Sheet 3*, complete column 3 in the table to calculate men's buying power for a *year's* worth of groceries, based on men's median earnings (page 39).

b. On *Activity Sheet 3*, graph the year and the percent of income used to buy groceries.

c. What assumption did you make to complete the table?

d. How has the buying power related to groceries changed since 1981? Is the rate of change constant?

e. What happens if you use a straight line to show the relationship between time and buying power for groceries?

3. Refer to the car prices data in Lesson 2, page 22.

a. Compare the change in a man's buying power for a new Honda Civic with the change in a man's buying power for groceries from 1981 to 1991.

b. Describe what the graph over time of a man's buying power in purchasing a new Honda might look like.

4. Use the cost per week of groceries from the table above.

a. Make a table and a graph of women's buying power.

b. How does women's buying power in purchasing groceries compare with that of men?

c. If women earn the median women's income, what do the results from a and **b** mean to families headed by women?

Unit II

Lines and Scatter Plots

EXPLORATORY LESSON

Balloons and More Balloons

Will a balloon that is blown up with lots of air stay in the air longer than a blown-up balloon that is smaller?

To answer the question, you are going to do an experiment. Collect the materials from your teacher, then work in small groups to design an experiment to collect data that might help you answer the question.

OBJECTIVE

Investigate the relationship between two variables.

EXPLORE

Balloons and More Balloons

Materials 4 or 5 balloons, string, scissors, a tape measure or meterstick, straws, stopwatch

Data Collection and Analysis

1. Plan the experiment.
   a. What is your hypothesis for the outcome of this experiment?
   b. What are the variables in your experiment?
   c. Describe in detail how you plan to conduct your experiment.
2. Conduct the experiment you designed, and record your data on *Activity Sheet 4*.
3. Make at least one scatter plot of the data collected by your group.
   a. Which variables in your experiment might have affected the results? Explain.

b. Are there any outliers in your data? If yes, why do you think the outlier(s) occurred?

c. Using what you learned in Unit I, summarize your results.

4. Predict the length of time a balloon with a circumference of 22 centimeters will stay in the air. Based on the data you have gathered and the analysis, how reliable do you think your prediction will be? Explain.

LESSON 5

Scatter Plots

When you were born, you were probably weighed and had your height measured. Why would doctors care about babies' heights, weights, and growth rates?

Is there some regularity in the growth rate of children?

Do babies' heights and weights also follow a growth pattern?

In Unit I, you studied scatter plots of data over time by investigating rates of change and whether or not the data were linear. For solving problems in which time is not a variable, but the data are still ordered pairs, scatter plots of the data can provide useful information. In some scatter plots, two or more points can appear with the same x-value and points can even be repeated or plotted on top of each other.

OBJECTIVES

Read and interpret a scatter plot. Summarize a linear relationship between two variables in a scatter plot by drawing a line.

INVESTIGATE

Baby Records

Parents and doctors keep records of a child's immunizations and height and weight, usually from the child's birth through high-school admission. The immunization records are required in many states for entrance into elementary school. Patterns in height and weight changes can be used as health indicators.

Discussion and Practice

1. Sophia, Sarah, and Tim all belong to the same family. The three tables on the next page show the record of each child's growth.

Height and Weight

Sophia

Age	Height	Weight
Birth	21.5 in.	9 lb 10 oz
3 mo	26.5	12 lb
6 mo	27	18 lb 10 oz
9 mo	29.5	21 lb
12 mo	30	23 lb
15 mo	33	25 lb
18 mo	34.5	26 lb 8 oz
21 mo	35	27 lb 10 oz
24 mo	35.5	28 lb 14 oz
27 mo	37	29 lb
30 mo	37.5	29 lb 5 oz
33 mo	38	30 lb 2 oz
36 mo	39.5	31 lb

Sarah

Height	Weight
22.5 in.	9 lb 2 oz
24.25	13 lb 4 oz
25.5	14 lb 4 oz
28.5	18 lb 9 oz
30.5	21 lb
31	22 lb 2 oz
32.5	23 lb 10 oz
33	24 lb 2 oz
35	24 lb 15 oz
35.5	25 lb. 3 oz
36	26 lb 9 oz
36.5	27 lb 14 oz
37.5	29 lb

Tim

Height	Weight
21 in.	7 lb 8 oz
24.5	14 lb 4 oz
28	18 lb
28.5	19 lb
31	19 lb 12 oz
32	21 lb 7 oz
33	23 lb
34	24 lb
35	26 lb
36	28 lb
37	29 lb
38	29 lb 10 oz
39	30 lb 5 oz

a. Have each member of your group choose one of the children and, plot the data (*age, height*) for the first 36 months of that child's life.

b. Is there a difference between Tim's height and the heights of his sisters? If so, how can you tell?

c. Describe how each child's height changed over time.

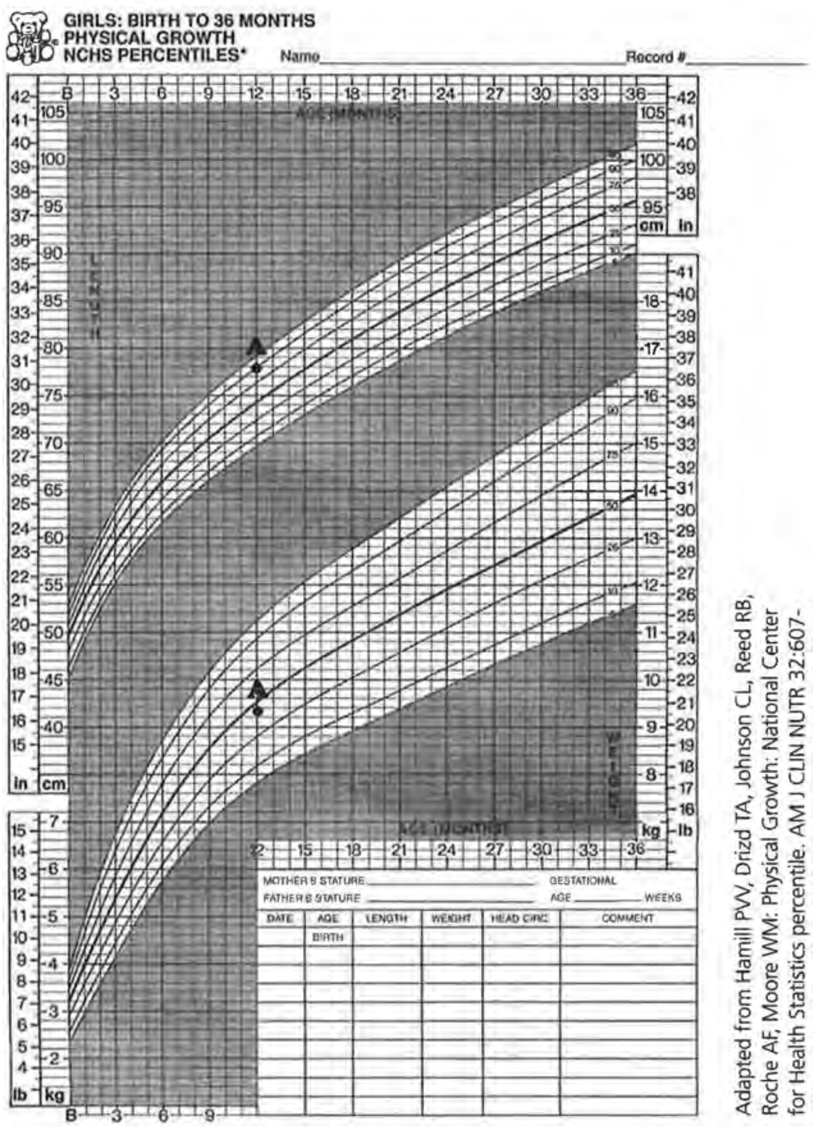
d. What do you think the typical height for a child in this family will be at age 2 years?

e. Sketch a graph that you think will show Tim's height over his lifetime.

Doctors actually use a very complicated chart with the three variables of age, weight, and height to show a child's growth. The vertical scale on this chart shows both weight and height while the horizontal scale shows time, or age. As a result there is (*time, height*) and (*time, weight*) on the same plot. The upper band on the chart shows a typical interval for a child's height at a given age; the lower band shows a typical interval for a child's weight at a given age. The curves inside the bands indicate the median height and weight for a child at a given age. From the growth chart, you can find the median heights and weights for given ages and intervals that suggest typical variability.

2. The following chart is used to record the growth of infant girls.

Girls' Growth Chart



Adapted from Hamill PVV, Drizd TA, Johnson CL, Reed RB, Roche AF, Moore WM: Physical Growth: National Center for Health Statistics percentile. AM J CLIN NUTR 32:607-629, 1979. Data from the Fels Longitudinal Study, Wright State University School of Medicine, Yellow Springs, Ohio. © 1982 Ross Products Division, Abbott Laboratories

a. If the two points labeled *A* on the growth chart represent Sarah, what can you say about her age, height, and weight?

b. What is the median height and weight for a typical 2-year-old girl?

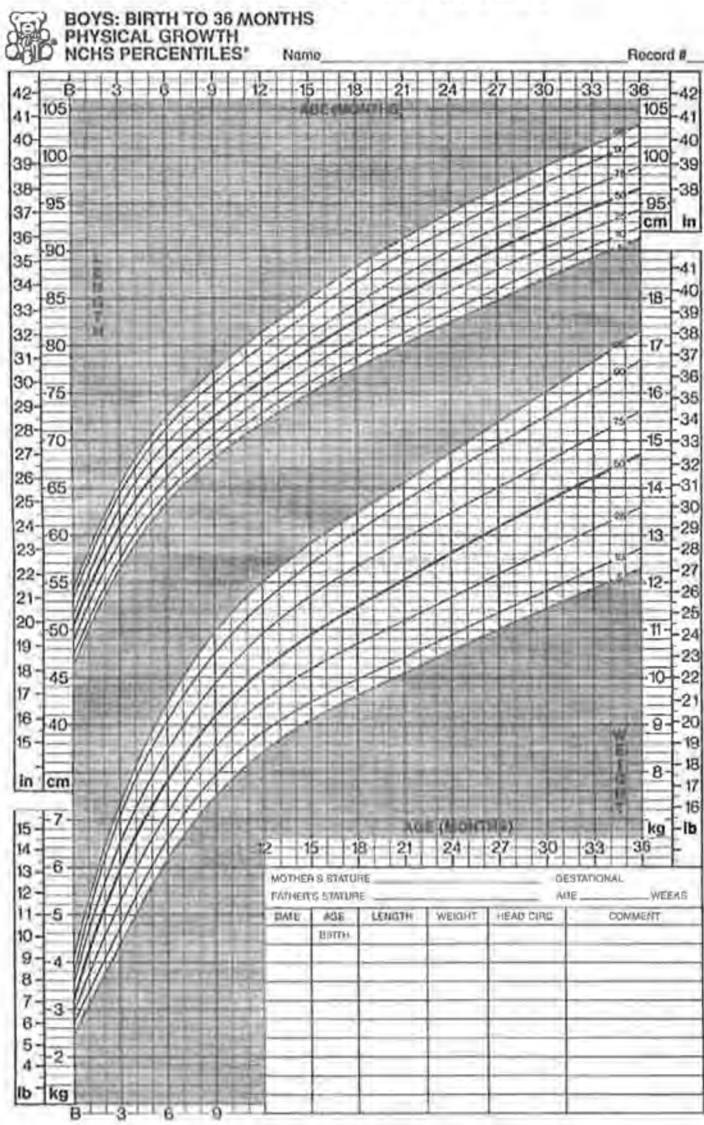
c. How might you expect the height and weight of a 2-year-old girl to vary?

d. Is the change in height constant as an infant girl grows? How can you tell? Is the change in weight constant? How can you tell?

e. Sophia was 34.5 inches tall and weighed 26 pounds 8 ounces at 18 months. Was her mother justified in claiming that she was very large for her age? Why or why not?

3. The following chart is used to record the growth of infant boys.

Boys' Growth Chart



Adapted from Hamill PVV, Drizd TA, Johnson CL, Reed RB, Roche AF, Moore WM: Physical Growth: National Center for Health Statistics percentile. AM J CLIN NUTR 32:607-629, 1979. Data from the Fels Longitudinal Study, Wright State University School of Medicine, Yellow Springs, Ohio.

a. How does the growth chart for infant girls compare with that for infant boys?

b. Have each member of your group now plot either Sarah's, Sophia's, or Tim's growth data on the appropriate growth charts shown on *Activity Sheet 5* or 6. Describe the growth of the child you chose.

4. You just investigated the relationship between time and height and also between time and weight. Now consider the relationship between height and weight.

a. Plot the data (*height, weight*) for the child you chose in Question 3 and draw a line to describe the relationship between the child's height and weight.

b. Write the equation of your line. What does the slope mean?

c. Compare your scatter plot and equation with the scatter plots and equations for the two other children by looking at the slopes and the overall patterns. Describe what you see.

d. Give two ways in which the growth patterns of the children were alike. Give two ways in which the growth patterns were different.

Whenever you make a scatter plot, you must first decide which variable to assign to the horizontal axis. Sometimes, the variable on the horizontal axis is the ***independent,*** or ***explanatory, variable;*** and the variable on the vertical axis is the ***dependent,*** or ***response,*** variable. In other words, given an ordered pair, the first variable (on the horizontal axis) can be used to predict or determine the second variable (on the vertical axis). In the scatter plots over time (*year, price*) from Unit I, the price of a car, for example, is determined by year, or the price is a ***function*** of time. In other cases, there is no clear explanatory (independent) and response (dependent) relationship; and therefore it is not important which variable is assigned to either the horizontal axis or to the vertical axis.

The heights and weights of a sample of high school football players are shown in the table on the next page. The relationship between the heights and weights of the football players can be analyzed.

High School Football Players

Height (inches)	Weight (pounds)	Position
70	160	QB
68	145	K
70	185	LB
74	190	FS
70	169	LB
71	155	K
68	140	FS
68	160	DB
62	125	SS
70	160	SS
67	154	RB
70	155	FS
71	150	LB
73	178	FS
72	190	LB
71	160	LB
72	205	LB
73	170	C
75	190	LB
76	230	DT
76	265	DT

(FS = Free Safety, RB = Running Back, DT = Defensive Tackle, SS = Strong Safety, K = Kicker, C = Center, LB = Linebacker, QB = Quarterback, DF = Defensive Back)

5. Look at the data in the table above.

a. Make a conjecture about the relationship between the heights and weights of the football players.

b. Use the grid on *Activity Sheet 7* to plot the data (*height, weight*). Based on your scatter plot, is your conjecture valid? Is it easier to see the relationship between height and weight when using the table or your scatter plot? Why?

c. Are there any outliers in the data? If so, describe them.

d. Draw a line on your scatter plot, write its equation, and use your equation to predict the weight of a football player whose height is 5 feet 11 inches.

SUMMARY

A ***scatter plot*** is a plot of a relationship using ordered pairs of data. There can be several points with the same x-value. If the pattern in a scatter plot is linear, you can draw a line and write its equation. If there is a variable you would like to predict, that variable is usually plotted on the vertical or y-axis.

Practice and Applications

6. A group of high school students collected data on the effect of age on a person's heart rate. Following are two tables: The first shows the data collected from people ages 5 to 20; the second shows data collected from people ages 21 to 50.

Age and Heart Rates

Ages 5 to 20

Age	Beats Per Minute	Age	Beats Per Minute
5	104	15	74
6	100	16	75
6	92	16	76
7	90	16	79
8	88	17	74
9	87	17	77
9	84	18	72
10	83	18	78
11	86	19	71
12	81	19	75
13	80	20	69
14	76	20	70

Ages 21 to 50

Age	Beats Per Minute
21	72
21	70
24	72
27	70
32	72
32	71
34	72
35	73
39	74
40	72
42	72
43	72
50	71
50	73

a. Using the same grid, make a scatter plot of *(age, beats per minute)* for each set of data. Use one symbol or color for the data in the 5 to 20 age group and a different symbol or color for the data in the 21 to 50 age group. Describe any trends you observe.

b. For which age group does there seem to be the most variability? How can you tell from the scatter plot?

c. At what age does a person's pulse rate seem to stabilize? How can you tell this from the scatter plot?

7. Think of a situation in which the data are ordered pairs. Collect at least 15 ordered pairs and make a scatter plot of your data. Using your scatter plot, what observations can you make about the situation or relationship in the data you collected?

LESSON 6

Graphs and Their Special Properties

Will every line cross an axis?

What do you know about a line when you know its rate of change?

What do you know about a linear equation when you know its rate of change?

OBJECTIVES

Recognize different forms of equations that represent a line and relate these forms to the graph. Find and interpret *x*- and *y*-intercepts of an equation and the zeros of an expression.

To summarize linear data, lines can be drawn in plots over time and in scatter plots. Sometimes it is difficult to tell, however, what a graph represents; two graphs can represent the same information and yet look totally different because of differences in the scales along the axes. Two equations can represent the same line but look different because they are written in different forms. What impressions can you get by looking at different representations, either in graphs or in equations? Are there any key features or common characteristics that are important?

INVESTIGATE

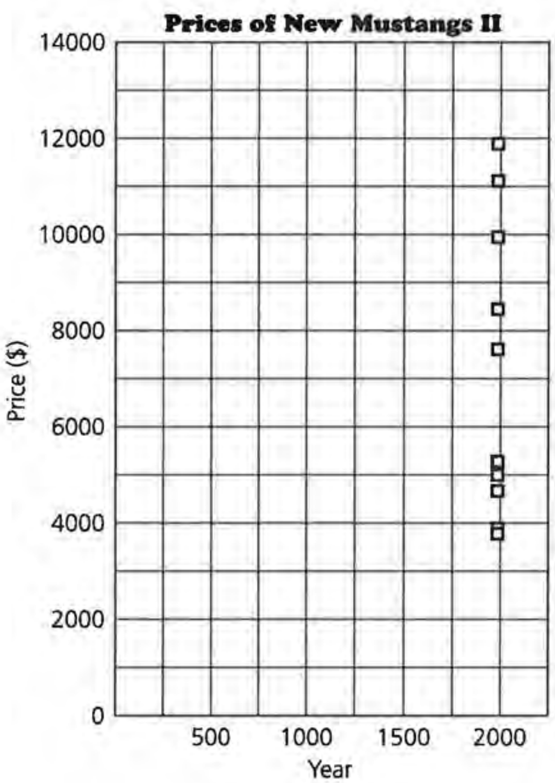
Cost of a Mustang

Sometimes the scales on the axes show only the part of a grid where the data are given or where the data would make sense.

Discussion and Practice

1. The following two scatter plots show the prices of a new Mustang taken from the data at the beginning of Lesson 2.

Source: data from *Kelly Blue Book,* 1994



a. Study the two scatter plots. Which plot shows more information about the relationship between the year and the price of a new Mustang? Why?

b. What horizontal scale will make sense when you are graphing the prices of cars? Write an expression that describes this axis.

c. A line has been drawn through the data in the following scatter plot. Write an equation of the line, relating price and year.

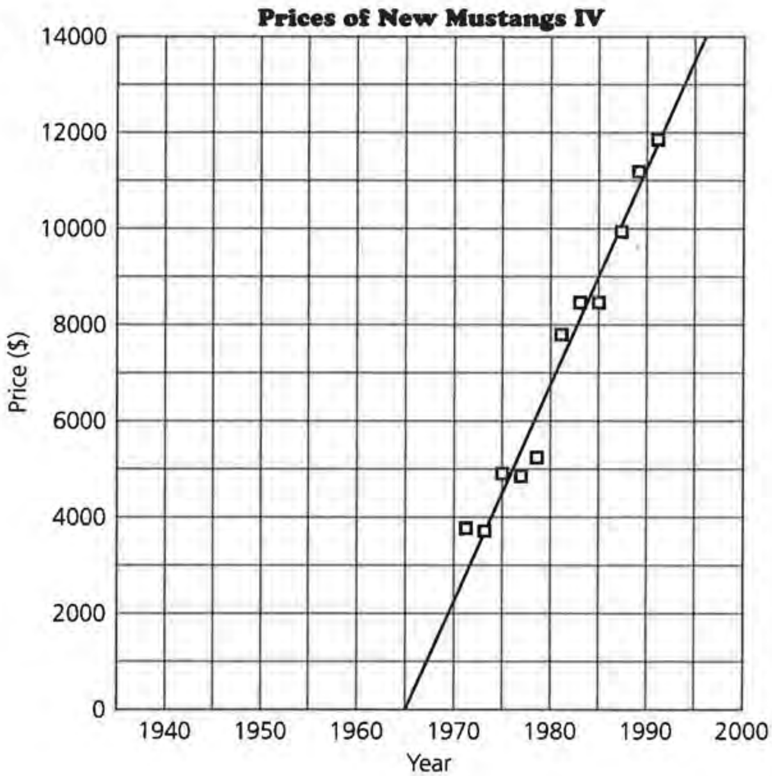
d. What is the ordered pair at C? What does this ordered pair represent?

Sometimes the points at which lines cross the x- and y-axes make sense for the data, and sometimes they do not.

2. Does the point at which the line crosses the horizontal axis in the preceding scatter plot make sense for the data? Why or why not?

3. Look at the next two scatter plots. Estimate the ordered pair where the line crosses the horizontal axis in each scatter plot. Describe what the ordered pairs would represent in each case. Does the ordered pair seem reasonable for each situation? Explain.

a.



Prices of New Mustangs IV
14000
12000
10000
8000
6000
4000
2000
0
Price ($)
1940
1950
1960
1970
1980
1990
2000
Year

b.

Long Playing (LP) Records Shipped
250
225
200
175
150
125
100
75
50
25
0
Number (millions) of LPs/EPs Shipped
1982
1984
1986
1988
1990
1992
1994
1996
Year

The scales and the relationship of the data to the axes are very important for interpreting the data. The range of values for x and y determines the part of the coordinate plane needed to plot the data. It can be called a viewing rectangle or window for the coordinate plane. For any ordered pair (x, y), the x-value is usually graphed along the horizontal axis (the horizontal coordinate), and the y-value is usually graphed along the vertical axis (the vertical coordinate).

4. There are two points that a line may or may not have—the points where the line crosses the axes. The ***x-intercept,*** the coordinate where a line crosses the x-axis, is often referred to as a ***zero.*** The ***y-intercept*** is the coordinate of the point where a line crosses the y-axis.

   a. Will a straight line always have x- and y-intercepts? Explain your answer.

   b. Why do you think the x-intercept of an equation is called a zero?

   c. Does it make sense to call the y-intercept a zero? Why or why not?

Suppose you are given an equation and asked to find the y-intercept and the x-intercept, or the zero. Just as you have used tables, graphs, and equations to tell whether or not two equations represent the same line, you can also use these same tables, graphs, and equations to identify the intercepts.

5. Consider the graph of a linear equation.

   a. How can you find the zero and y-intercept of a linear equation from its graph?

   b. How can you find the zero and y-intercept of a linear equation by studying a set of ordered pairs that satisfy the equation?

6. Consider the equation $P = 10{,}000 - 500(A - 2)$ where A represents the age of a used car and P represents the price.

   a. How much will a 5-year-old car cost?

   b. If $A = 0$, find P. What does the point $(0, P)$ represent when describing used cars?

   c. If $P = 0$, find A. What does the point $(A, 0)$ represent when describing used cars?

7. Use the equation for (*age, price*) from Question 6 above.

a. What are some values for A and P that will make sense in the problem? Use those values to select a viewing window, and sketch a graph of the line.

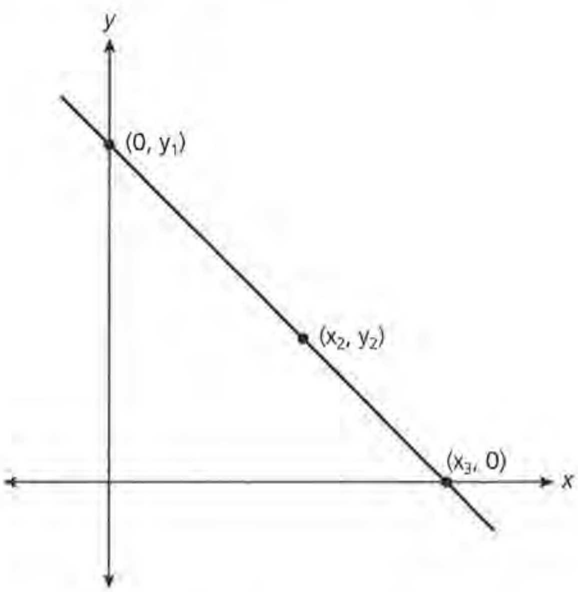
b. What is the slope of the line, and what does it represent?

c. Find the zero from the graph, and compare it to the zero you found in Question 6. Use the point and the slope to write the equation of the line.

d. Find the y-intercept and compare it to the y-intercept you found in Question 6. Use that point and the slope, to write another equation of the line.

e. Prove that the equations from Question 6 and **c** and **d** of this problem are all equivalent.

Consider the following graph of a line showing three labeled points and a slope m.



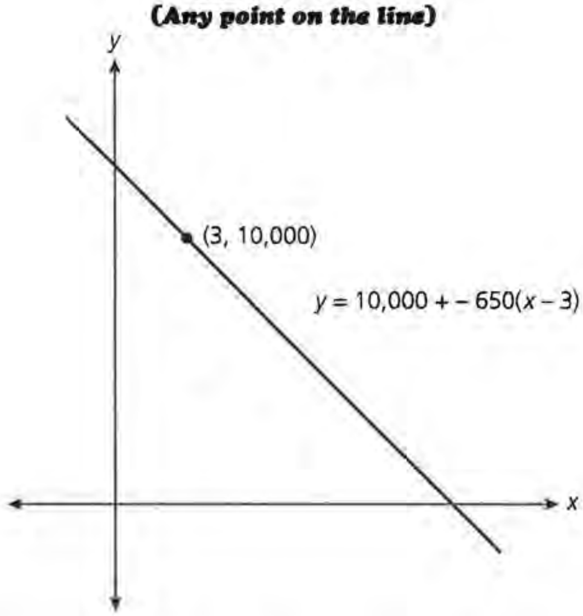
Any specific point (x_2, y_2) on the line can be used, along with the rate of change, m, to write the equation of the line:

$$y = y_2 + m(x - x_2)$$

For example, this general equation can be used to describe the following situation: Suppose a 3-year-old car is worth \$10,000. If the rate of change in price is a decrease of \$650 per year, using (3, 10,000), then the equation of the line is

$y = \$10{,}000 + -650(x - 3)$ as shown in the following graph.

Used Car Prices
(Any point on the line)



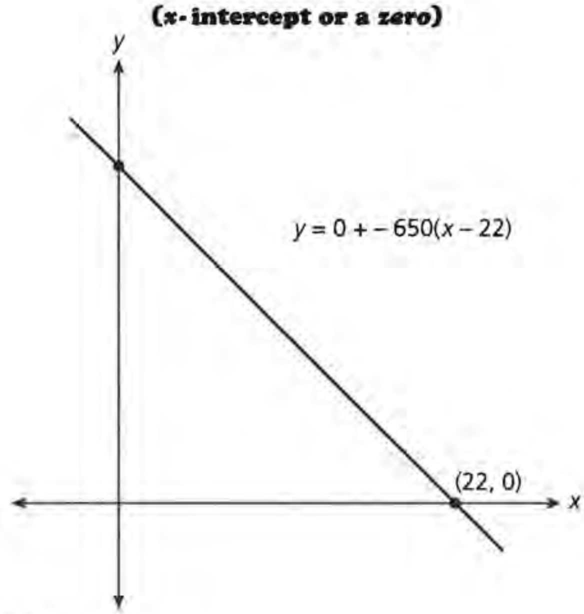
Suppose the point $(x_3, 0)$ is where the line crosses the horizontal axis. Then you can use this point (the x-intercept or zero), along with the slope, m, to write the equation of the line:

$y = 0 + m(x - x_3)$, or

$y = m(x - x_3)$

In the context of the example about used cars, the x-intercept or zero is (22, 0). This is equivalent to saying that the used car is worth nothing when it is 22 years old. The equation of the line is $y = 0 + {}^{-}650(x - 22)$, or $y = {}^{-}650(x - 22)$ as shown on the following graph.

Used Car Prices
(x- intercept or a zero)

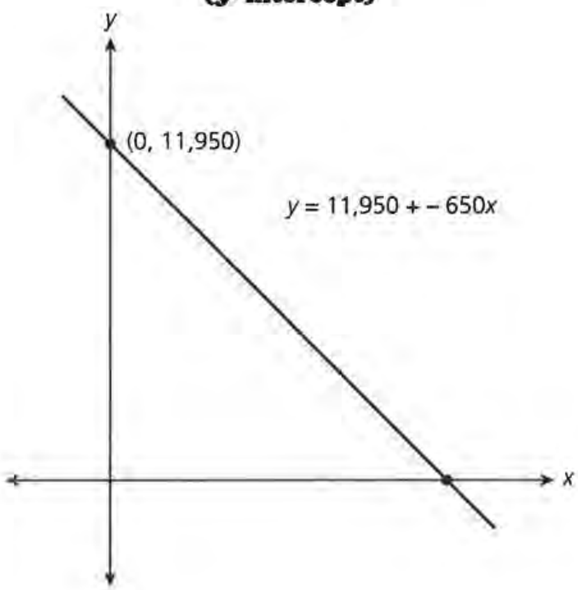


Suppose the point $(0, y_1)$ is where the line crosses the vertical axis. Then you could use this point (the y-intercept) to write the equation of the line:

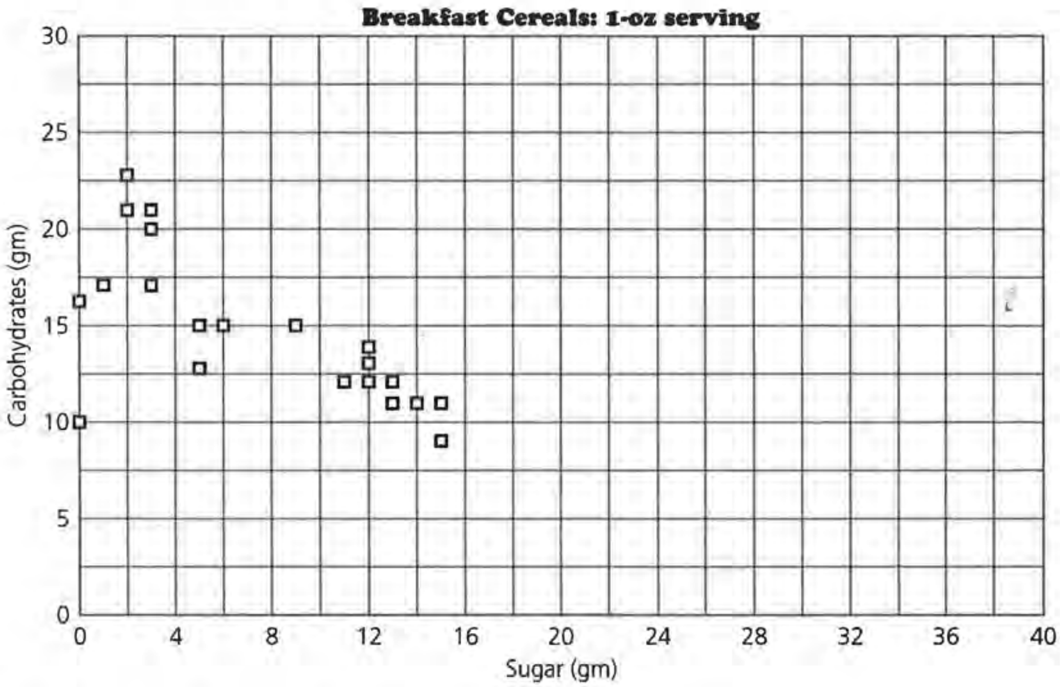
$$y = y_1 + m(x - 0), \text{ or } y = y_1 + mx$$

In the context of the used car example, the y-intercept is (0, 11,950). This is equivalent to saying that when a car is brand new (zero years old), it is worth \$11,950. The equation of the line is $y = 11{,}950 + {}^{-}650(x - 0)$, or $y = 11{,}950 + {}^{-}650x$ as shown on the graph below.

Used Car Prices (y-intercept)

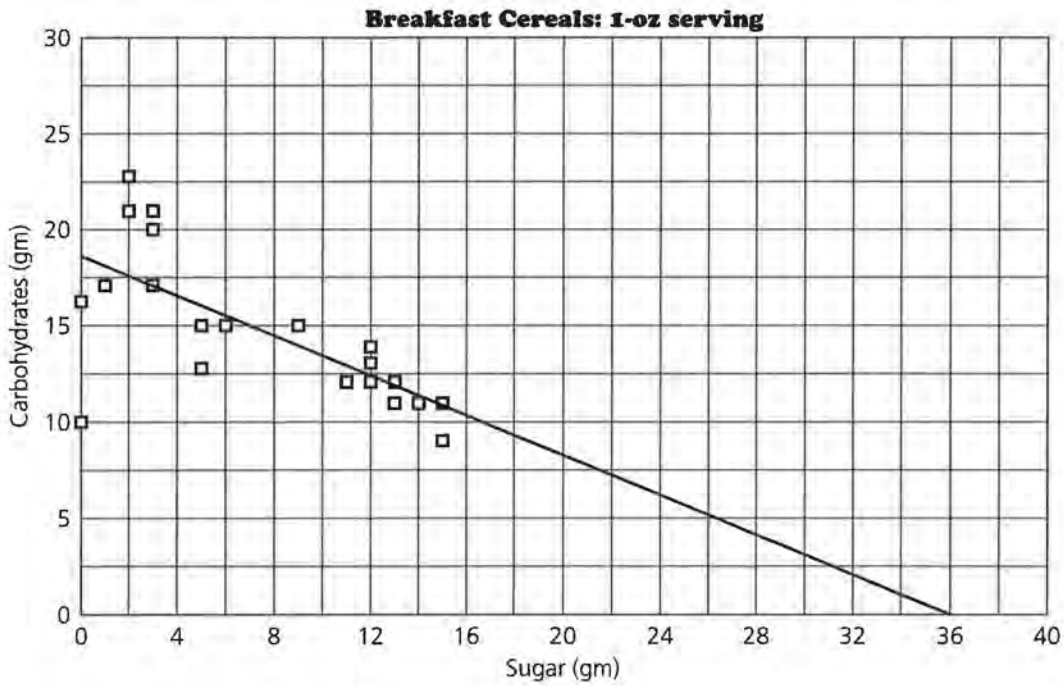


8. The following scatter plot shows the relationship between the grams of sugar and the grams of carbohydrates in a 1-ounce (oz) serving of various breakfast cereals.



a. In general, describe the relationship between the amount of sugar and carbohydrates in the cereals.

b. A line has been drawn on the following scatter plot to describe the relationship. Estimate the zero and the y-intercept from the graph. What does each represent with regard to the amount of sugar and the amount of carbohydrates in the breakfast cereals?



c. Write the equation of the line.

d. Use your equation to find the zero. How does this value compare with the estimate you made in **b**? How does the zero relate to the data?

9. What is the relationship between the television viewing habits of men and of women? Do they watch the same programs? The chart on the next page shows the percentage of men and women who watched the top 30 TV programs during the 1992–1993 television season according to Nielsen Media Research.

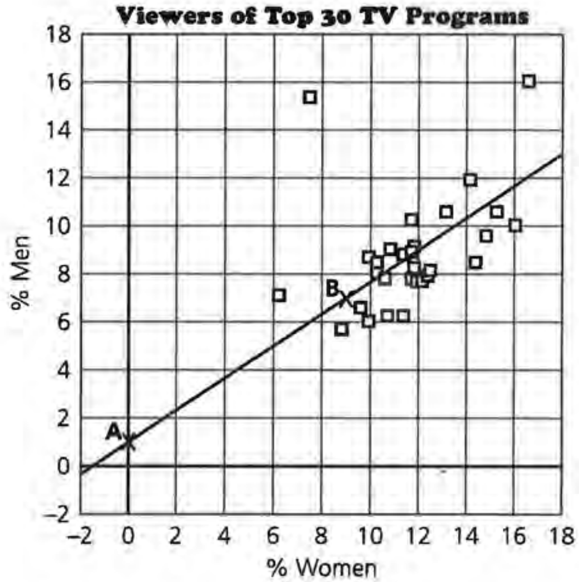
Television Viewing Habits

Rank	Program	% Women	% Men
1	*60 Minutes*	16.5	16.1
2	*Roseanne*	15.3	10.6
3	*Home Improvement*	14.1	12.0
4	*Murphy Brown*	14.7	9.6
5	*Murder She Wrote*	16.0	10.1
6	*Coach*	13.1	10.7
7	*NFL Monday Night Football*	7.6	15.3
8	*CBS Sunday Night Movie*	14.3	8.4
9	*Cheers*	11.6	10.4
10	*Full House*	10.7	6.4
11	*Northern Exposure*	11.8	9.0
12	*Rescue 911*	11.7	9.2
12	*20/20*	11.6	9.0
14	*CBS Tuesday Night Movie*	11.8	8.3
15	*Love and War*	11.9	7.7
16	*Prince of Bel-Air*	9.5	6.6
16	*Hangin' with Mr. Cooper*	9.9	6.1
16	*Jackie Thomas Show*	10.5	7.8
19	*Evening Shades*	12.4	8.2
20	*Hearts Afire*	12.1	7.7
20	*Unsolved Mysteries*	11.2	8.8
22	*Primetime Live*	10.7	9.1
23	*NBC Monday Night Movie*	11.2	6.4
24	*Dr. Quinn, Medicine Woman*	12.3	8.0
25	*Seinfeld*	9.9	8.7
26	*Blossom*	8.7	5.6
26	*48 Hours*	10.2	8.4
28	*ABC Sunday Night Movie*	10.2	8.2
29	*Matlock*	11.6	7.8
30	*Simpsons*	6.3	7.2
30	*Wings*	9.5	8.1

Source: data from *The World Almanac and Book of Facts,* 1994.

a. Why do you think there were no TV programs that ranked 13th, 17th, and 18th on the chart?

b. With the exception of at least one program, the pattern of the data appears to be linear as shown on the following scatter plot. Name the program(s), and tell why you think it is an outlier(s).



c. Write a description in words or symbols of the range of values that make sense for the scale on each axis.

d. What is the y-intercept and the zero or x-intercept of the line? Do these both make sense in terms of the data? Explain.

10. Refer to the table *Television Viewing Habits.*

a. Find the rate of change, and use point A to write an equation of the line relating the percentage of men to the percentage of women who watched the top 30 TV programs.

b. Use point B and your rate of change to write an equation of the line.

c. Show that the two equations are equivalent.

d. If you knew that approximately 10% of the women watched a given TV program, about how many men would you predict watched the same program?

11. Draw the line $y = x$ on *Activity Sheet 8.*

a. Suppose a data point lies on this line. What does this data point represent? What does a data point above the line represent? What does a data point below the line represent?

b. What does this line tell you about the percentage of men and the percentage of women who watched the TV programs?

c. Suppose you used the line $y = x$ to predict the TV viewing habits of men and women. How is this different from using the equation in Question 10?

12. Recall that the form of an equation can help you find the y-intercept and zero.

a. Write the three general forms you can use to write a linear equation.

b. Which of the forms use the zero as the starting point?

13. Consider $y = 15 - 5(x - 4)$.

a. Suppose (4, 15) is a point on the line. How can you verify this by using the equation?

b. If you use the Distributive Property and combine terms, then you can write the following equation:

$y = 15 - 5(x - 4)$

$y = 15 - 5x + 20$

$y = 35 - 5x$

Is $y = 35 - 5x$ equivalent to $y = 10(x - 4)$? Why or why not?

c. What is the slope of $y = 35 - 5x$? The y-intercept? How can you find them from this equation?

d. Given the equation $y = 20 + 3x$, what is the y-intercept?

The Distributive Property allows you to distribute a common multiplier to a sum or to write a sum as a product of a common factor and an expression. If the slope is a common factor of both terms, you can distribute the slope from the sum.

$y = 35 + {}^-5x$

$y = {}^-5({}^-7) + {}^-5(x)$

$y = {}^-5({}^-7 + x)$

$y = {}^-5(x + {}^-7)$

So, $x = 7$ is the *zero* of the equation.

14. Think about the *zero* of a linear equation.

a. When $x = 7$, what is the value of y in the equation $y = {}^-5(x + {}^-7)$?

b. Why is $x = 7$ called the zero?

c. Jenny thinks the zero for $y = 10 + 5(x - 2)$ is $x = 2$. Is she right? Why or why not?

15. Find the zeros of each of the following.

a. $y = 6(x - 4)$ and $y = 6(4 - x)$

b. Compare the lines represented by the equations in a.

c. Is $6(x - 4) = 6(4 - x)$ true or false? Verify your answer.

16. For each of the following equations, identify the zeros and the y-intercepts. Indicate how you found your answer.

a. $y = 10(x - 5)$

b. $y = 1000 + {}^{-}20x$

c. $y = 6(8 - x)$

d. $y = 100x$

SUMMARY

To understand the relationship between a line and its graph, it is helpful to consider several characteristics.

- To make a scatter plot from the data, choose appropriate scales including maximum and minimum values for both axes.
- To make a graph from an equation, choose an appropriate viewing window.
- To choose an appropriate window, inspect the equation for a reasonable starting point and ending point, or generate a set of ordered pairs that satisfies the equation.

Sometimes, two special points can be useful.

- For the y-intercept at $(0, y)$, you can find the coordinates of this point by doing the following: substituting 0 for x in the equation; reading the point where the graph crosses the y-axis; or writing the equation in the form $y = y_1 + mx$ and recognizing that y_1 is the y-intercept.
- For the x-intercept at $(x, 0)$, you can find the coordinates of this point by doing the following: substituting 0 for y in the equation; reading the point where the graph crosses the x- axis; or writing the equation as a product in factored form and finding the value for x that makes the factor 0.

- Sometimes the intercepts make sense for the data; often they do not. For a given slope m, understanding the relation between the starting point used to write the equation and the form of the equation can be useful. If m is the slope:

  If (x_1, y_1) is any point on the line, then $y = y_1 + m(x - x_1)$.

  If $(0, y_2)$ is the y-intercept, then $y = y_2 + mx$.

  If $(x_3, 0)$ is the zero, then $y = m(x - x_3)$.

Practice and Applications

17. Find the zero and the y-intercept for each of the equations.

a. $y = 20 + 15x$

b. $n = 200 - 10(x - 1970)$

c. $a = 6 + 0.2(x - 15)$

d. $t = 5(x - 6)$

e. What viewing window or set of scales for the axes would make sense for each of the equations above?

18. Use the Distributive Property to rewrite each of the expressions as a product with a common factor.

a. $10x - 30$

b. $15 + 5x$

c. $100 + 20x$

19. The following table shows the cost of used ¾-ton Chevrolet pickup trucks as a function of age.

Used Chevrolet Pickup Trucks

Year	Age (years)	Price ($)
1974	20	2,500
1976	18	2,200
1980	14	3,650
1981	13	2,000
1983	11	5,388
1986	8	6,900
1988	6	9,995
1989	5	9,300
1992	2	17,595
1992	2	17,500
1993	1	23,000

Source: *The Milwaukee Journal,* December 12, 1993.

a. Make a scatter plot (*age, price*) of the data. Draw a line that you think best represents the relationship in the data, and write the equation of that line.

b. Locate the point at which the line you drew intersects the vertical axis or y-axis. Write the equation in the form $y = y_1 + mx$. What do y_1 and m represent in the context of the data?

c. Estimate the point at which the line intersects the horizontal or x-axis. What does this point mean in the context of the data?

20. Write a description of appropriate scales for the axes or viewing window for each of the following situations. In each case, describe what the y-intercept and the zero would tell you about the data.

a. The number of free throws attempted and the number of free throws made by a basketball player

b. The cost of a camera over time

c. The percentage of rainy days and the percentage of sunny days for cities in the United States

d. The fats and calories in fast foods

21. Describe a real-world situation and an appropriate viewing window for each of the equations.

a. $n = 30 + 5(x - 1970)$

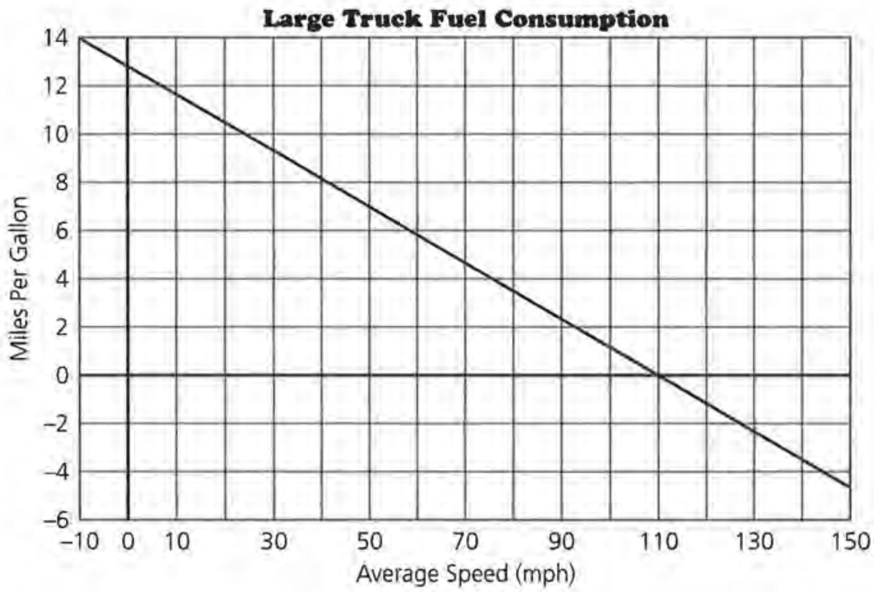
b. $t = 120 - 0.8x$

22. Refer to the graph below.

a. What does the graph represent?

b. What do you think would be a more appropriate range of values for speed? For miles per gallon?

c. Identify the zeros and the intercepts from the graph and tell what they might mean in the context of the data.



23. Determine an appropriate viewing window for each of the following:

a. To see the y-intercept and zero for $y = 50 - 0.5(x - 20)$

b. To see the line $c = 4.30 + 1.5(x - 1975)$ when $1960 \leq x \leq 2000$

c. To see the y-intercept and zero for $y = 20 + 10x$

d. To see the line $y = -110 + 4x$ for $150 \leq x \leq 175$

24. How well can you make a quick sketch of the graph of a linear equation? Work in partners on the problems in Sets A and B listed at the top of the next page.

a. Suppose you use the zero and y-intercept for each line. With your partner, discuss how to make a quick sketch of the line for each equation.

b. Sketch the line for each equation in Set A by hand while your partner uses a graphing calculator. Compare your results. Change roles and repeat the process using Set B.

Set A	Set B
$y = 10 - 5(x - 1970)$	$y = 210 + 0.5x$
$y = 50 + (1.10)x$	$y = 87 - \frac{3}{8}x$
$y = 50 + \frac{1}{10}(x - 0)$	$y = 81 - \frac{3}{8}(x - 16)$
$y = 12 + 5(x - 2)$	$y = 50 + 10(x - 0)$
$y = 2 + 5x$	$y = 50 + 10(x - 1950)$
$y = 65 + 0x$	$y = 80 + \frac{5}{7}(x - 5)$
$y = 1200 + 0.5(x - 1980)$	$y = 75 + \frac{5}{7}(x - 12)$

25. For each equation, find x when $y = 0$. What does the result tell you?

a. $y = 50 + 4(x - 2)$

b. $y = 17 - 0.2x$

26. The number of deaths (d) on highways for a given year (t) can be approximated by the equation $d = 47{,}000 + 1445(t - 1988)$.

a. What are the starting point and the slope? What do they tell you about the number of deaths on the highway?

b. Draw a sketch of this equation.

c. Estimate the number of highway deaths in 1991. How did you make your estimate?

d. Find the zero and interpret it in the context of the data. How did you find the zero?

27. The Presidential Physical Fitness Award Program is offered in some United States schools. For the Qualifying Standards, the relationship between a girl's age (16 and under) and the time (in minutes) it takes for a girl to run one mile can be estimated by the equation $T = 13.06 + {}^{-}0.36A$.

a. What is the slope of the line, and what does the slope tell you about the time it takes a girl to run a mile?

b. Find an appropriate window for the graph of the equation. Make a sketch of the line.

c. Find the y-intercept and the zero. Write a description of each in terms of the situation.

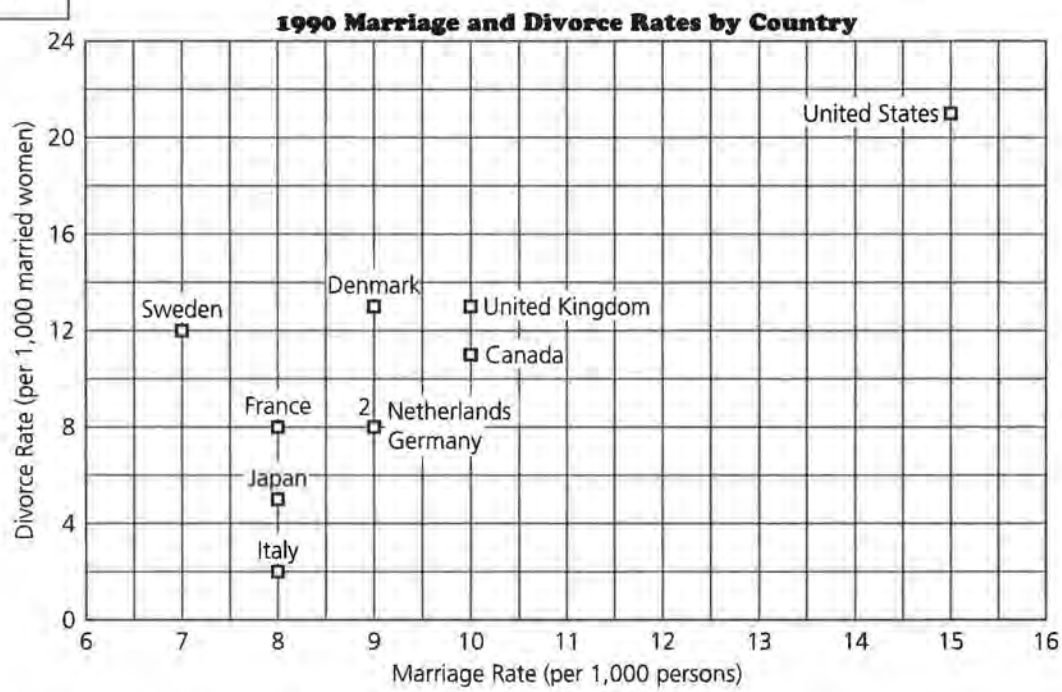
d. Why do you think the girls' ages are restricted to 16 and under?

Vital Statistics

OBJECTIVES

Use intercepts, rate of change, and lines to analyze a situation.

1. The 1990 marriage and divorce rates by country are shown in the following scatter plot. Refer to the scatter plot to answer each question.



Source: data from *American Almanac*, 1994–1995

a. In 1990, what was the marriage and divorce rate in Japan? Explain what the values indicate.

b. Which country had the greatest marriage rate and the greatest divorce rate? How can you tell?

c. Which country had the fewest divorces?

d. In 1990, how did the marriage and divorce rates in Denmark compare with those in the United Kingdom?

2. The following table shows the life expectancy at birth (in years), daily calorie supply for adults, and the infant mortality rate (per 1000 births) for selected countries in 1991. (Infant mortality is the number of deaths for children less than one year old, exclusive of fetal deaths.)

1991 Life Expectancies, Daily Calories, and Infant Mortality by Country

Country	Life Expectancy (years)	Daily Calorie Supply (for adults)	Infant Mortality (per 1000 births)
Ethiopia	48	1,667	193
Zaire	52	1,991	150
Argentina	71	3,113	34
Bolivia	60	1,916	115
Mexico	70	3,052	45
United States	76	3,671	11
China	70	2,639	33
Indonesia	62	2,750	78
United Kingdom	76	3,149	10
Hungary	71	3,644	19
Switzerland	78	3,562	8
Libya	63	3,324	88

Source: data from *The Universal Almanac,* 1994.

a. In 1991, which country had the highest infant mortality rate? Explain how you think the mortality rate affects life expectancy.

b. Plot the data (*infant mortality, life expectancy*). Then draw a line on your scatter plot to summarize the data, and write its equation. What does the slope of your line tell you about the data?

c. Find the zero and the y-intercept. Explain how you found each.

d. What do the zero and the y-intercept represent in terms of the data?

3. Suppose the equation $y = 285 + {}^{-}0.08x$ summarizes the relationship (*calories, infant mortality rate*) as shown in the table above.

a. Find each of the following: the zero, the slope, and the y-intercept, and then indicate what each represents in terms of the data.

b. Describe how you found your answers for **a**.

Lines on Scatter Plots

How do you know if the line relating the percentage of sugar to the amount of carbohydrates in cereals, or the line relating the cost of a car to its age, is a *good* line?

If you can draw several different lines to summarize the data, how do you know if one line is better than another?

OBJECTIVE

Find a *good* line for a scatter plot by considering a measure of error.

In earlier lessons, you investigated drawing a line to summarize a relationship between two variables. You learned to write an equation or rule for the line you drew and learned how to tell if different algebraic representations of your line were equivalent. Also, you studied some special characteristics of a line and that line's equation. Questions still remain, however. How good is your line? How well does it fit the data? The activities in this lesson will help you to answer these questions.

INVESTIGATE

Ages of the Rich and Famous

In 1997, Walt Disney World celebrated its 25th anniversary. Often people are surprised when they learn the ages of famous people or of things that are in the news or are part of their daily lives. This lesson will give you an opportunity to find out how well you can estimate the ages of some famous people.

Discussion and Practice

1. Here is a list of famous people. Make a table and estimate how old you think each person is. Then your teacher will give you the ages of these famous people.

Ages of Famous People

	Estimated Age (years)	Actual Age (years)
Nancy Reagan		
Mister Rogers		
Sandra Day O'Connor		
Macaulay Culkin		
Eddie Murphy		
Tom Brokaw		
Roseanne		
Ringo Starr		
Frank Sinatra		
Oprah Winfrey		
Jon Bon Jovi		
Garth Brooks		
Jane Fonda		

2. How well were you able to estimate the ages of these people? To help answer this question, make a scatter plot with your estimates of the ages on the horizontal axis and the actual ages on the vertical axis, (*estimated age, actual age*).

a. How do you think you should decide if someone is a good estimator of the ages of these famous people?

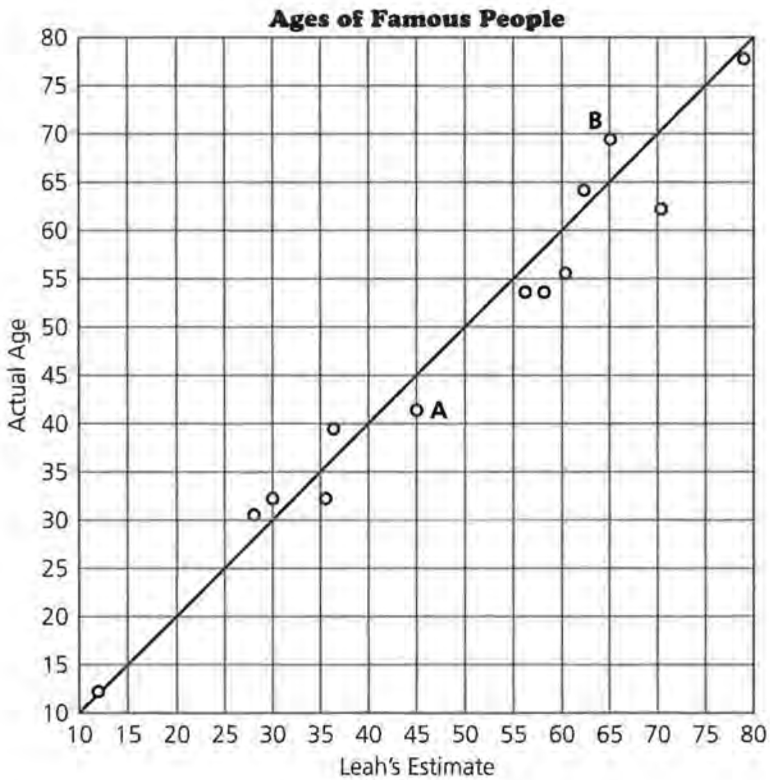
b. Where will a point lie if your estimate is correct? Draw a line on your scatter plot representing estimates that are 100% accurate. Write the equation of the line.

c. What does it mean if a point is above this line?

d. In general, are you an overestimator or an underestimator? How can you tell?

3. Look at the scatter plots of other students. Decide who among you is the best estimator, and give a reason why. Be prepared to explain your choice to the class.

4. Following is a scatter plot of Leah's estimates and the actual ages for another group of famous people.



a. What does A represent?

b. What does the vertical distance from A to the line represent?

c. Name the ordered pair that represents the person for whom Leah had the worst estimate. How can you find that ordered pair using the graph?

d. Suppose the difference between Leah's estimate and the actual age is called her "error" in estimating. What was Leah's error for the person represented by B? Describe how to find this error using her scatter plot. What is the difference between the error for B and the error for A?

5. Leah decided to find out how far off all of her estimates were, or to find her total error. She claimed that overall her total error was only 14 years, and she wrote $\sum$ errors = 14 ($\sum$ means "sum").

a. Find all of the differences between her estimated ages and the actual ages, and record them in a table.

b. What do you think Leah did to find $\sum$ errors = 14?

c. Do you think her method was a good one? Why or why not?

6. Use your estimates of age and the scatter plot you made in Question 2.

a. How far off was each of your estimates? What is the sum of your errors?

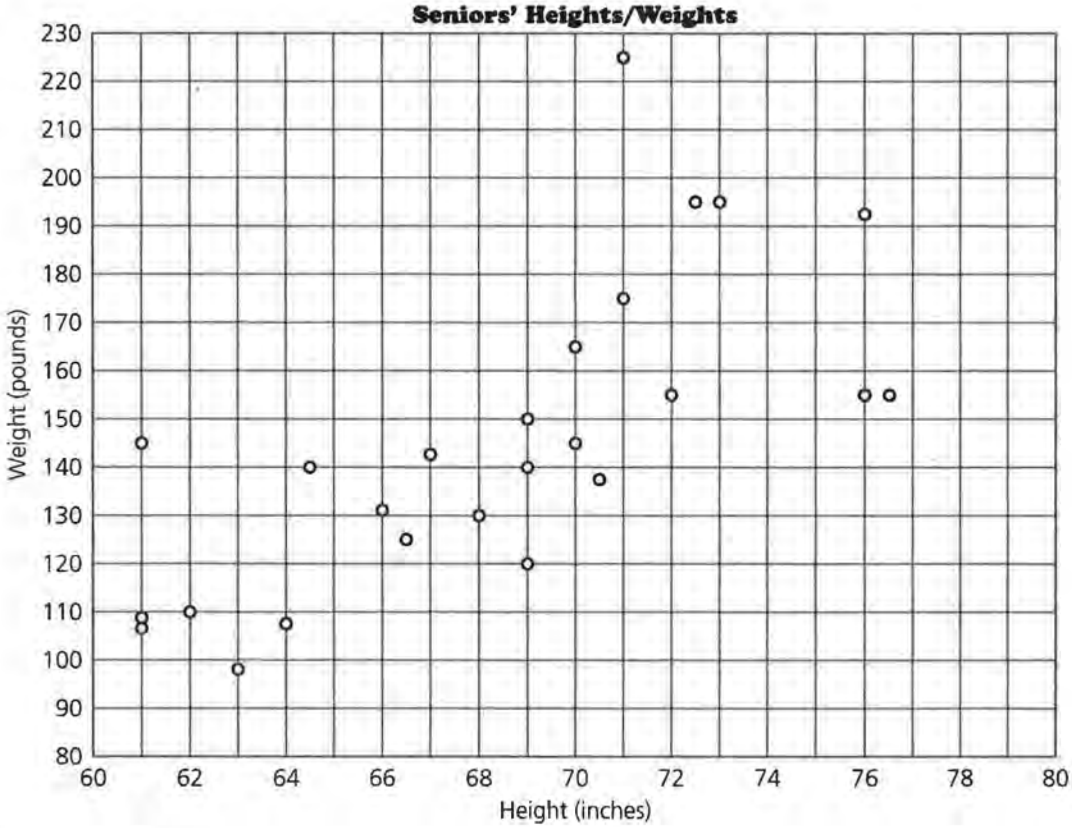
b. On average, how far off were you for each estimate? What does this tell you about your ability to estimate ages of this group of famous people?

c. If you define the "best" estimator as the person with the least error, who is the best estimator in your class?

In an earlier lesson, you investigated the relationship between the heights and weights of football players. Do you think the same relationship exists for students who are not football players? The following data show the heights and weights of 25 high school seniors, both male and female. A scatter plot of the data is shown on the next page.

High School Seniors' Heights/Weights

Males' Height (inches)	Weight (pounds)	Females' Height (inches)	Weight (pounds)
72.5	195	66	132
70.5	137	63	97
72	156	61	145
70	145	61	108
71	175	64	106
70	165	66.5	125
73	195	64.5	140
76	155	69	120
67	143	68	130
76.5	155	61	105
76	192	69	140
69	149	62	110
71	225		



7. Refer to the scatter plot above.

a. Describe the pattern you see in the seniors' heights and weights.

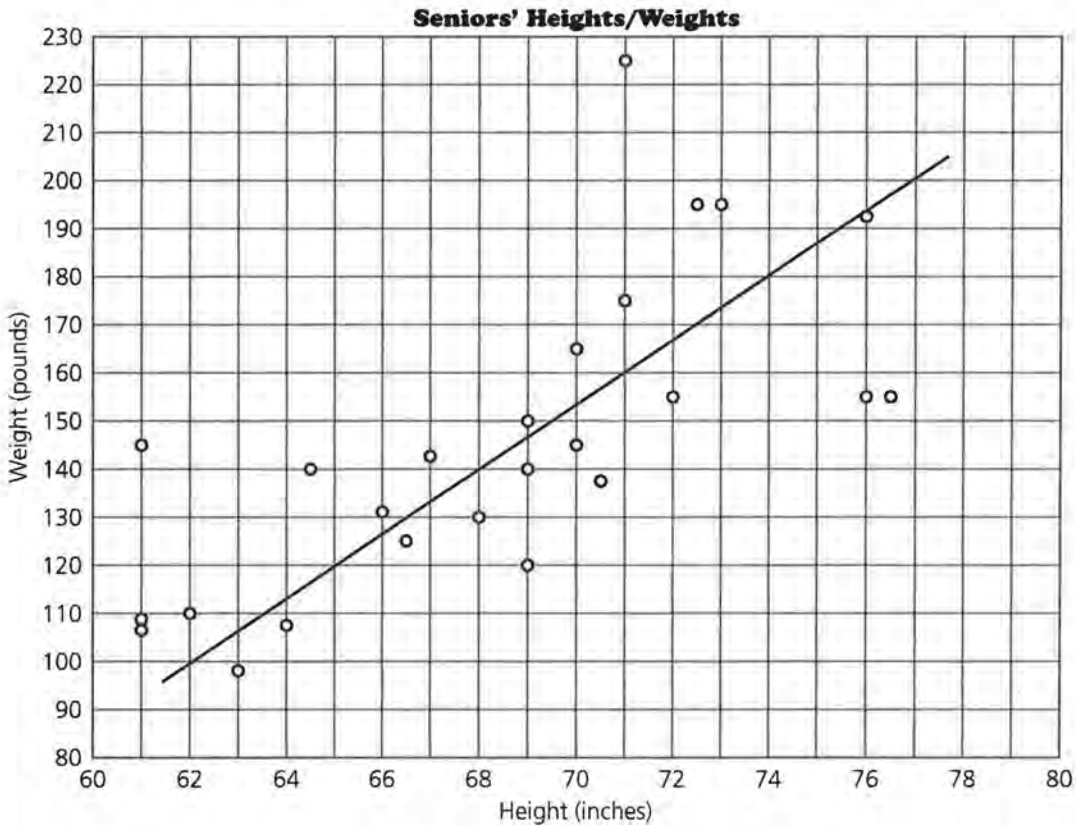
b. Use *Activity Sheet 9* to star the points representing the heights and weights of females. What observations can you make?

c. Draw a line on the scatter plot that you think will summarize or fit the data for all students, male and female. Find the equation of your line.

d. How does your equation for the heights and weights of a sample of seniors compare with the equation relating the heights and weights of football players in Lesson 5, Question 5?

8. Look at the line you drew in Question 7. Working with your partner, decide who has the "better" line. How did you decide?

9. Which line is the best for predicting a given height? One way to find out is to test your line using the data points you know. A line has been drawn on the scatter plot that follows. The point (68, 140) is on the line.



a. Mark this point on the scatter plot on *Activity Sheet 10.*

b. Look at the table of data to find the weight of the senior who is 68 inches tall. Use *Activity Sheet 10* to mark the point on the plot.

c. What is the difference between the weight that is predicted by the line drawn on the scatter plot and the actual weight from the data in the table?

10. Refer to Question 7 and *Activity Sheet 9.*

a. How well did your line predict the weight for a height of 72 inches?

b. Who in your group was the best predictor of weight for a height of 72 inches?

11. Refer to the scatter plot (*height, weight*) shown above with Question 9.

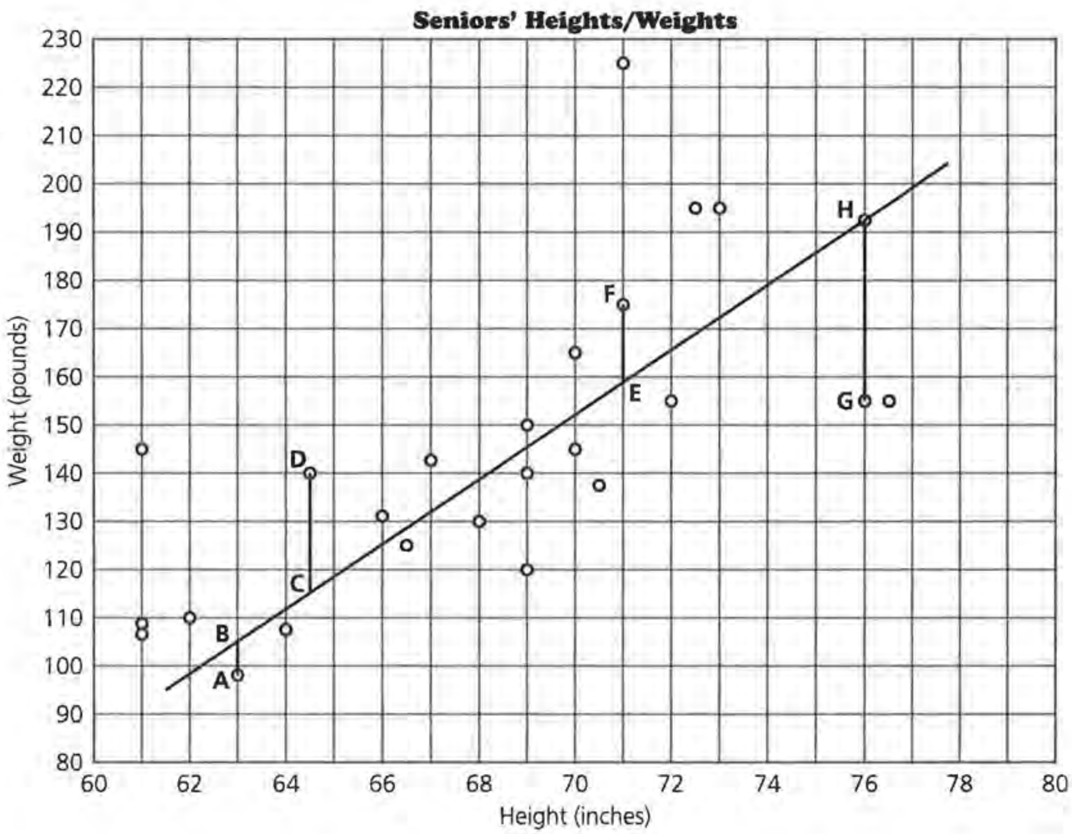
a. For which height(s) does the line predict the actual weight(s)?

b. Overall, does the line overestimate the weight or underestimate the weight? How can you tell?

c. For which point will the weight that is predicted from the line be farthest from the actual weight? How can you use the scatter plot to decide?

d. Find the point (62, 110). Describe how you can show on the scatter plot the difference between the actual and predicted weight.

12. Consider the line segments shown on the scatter plot below.



a. Using the scatter plot, estimate the length of each segment: *AB, CD, EF,* and *GH*. What does each length represent?

b. Suppose the equation of the line shown in the scatter plot is $W = 140 + 6.7(H - 68)$.What do the numbers in the equation represent in terms of the data?

c. Use this equation to predict the weight of a senior who is 66 inches tall. How much does this senior actually weigh? How close is the weight of this senior predicted by the equation to this senior's actual weight?

d. Use the data in the table to find the coordinates of *A, D, F,* and *G*. Then use the equation to determine the length of the segments AB, CD, EF, and GH, as shown in the scatter plot.

13. Recall that you can generate a table of values for any equation.

a. Use the equation for the line you drew in Question 7 to predict the weight for each height. Record your data in the table on *Activity Sheet 10*.

b. Find $y_i - y$, the difference between any given actual and predicted weight. What does a negative difference in weight represent?

c. How are the predicted weights shown on the scatter plot? Give two examples.

14. Use the table on *Activity Sheet 10*.

a. Find a method to determine the average error in your predictions of weight. Be prepared to explain your method.

b. Compare your error with the errors of your classmates. Who has the *least* error and thus, the *best* line?

15. To find an *average* error, you need to find the difference between the actual weight and the predicted weight. These differences are called ***residuals***.

a. Does it make sense to add the differences, or residuals, to find an average error? Why or why not?

b. If y_i is any actual value and y is any predicted value for a given height, what does $\Sigma(y_i - y)$ represent?

c. Recall the interval regarding the median height on the infant growth charts in Lesson 5, Questions 2 and 3. Using *Activity Sheet 10*, sketch on the scatter plot a band that you think shows the typical error in using your line to predict weight.

There are two ways to avoid the problem of negative residuals canceling positive residuals. Use the absolute value of each residual or use the square of each residual before adding them. Squaring is more typical, but taking the absolute value may seem more logical. You can estimate your error in making a prediction by finding an "average" error:

- Find the differences between your estimate and the actual value.
- Square or take the absolute values of the differences.
- Find the sum of the squared or absolute differences.
- Find the average by dividing the sum by the number of differences you have.
- If you use the squared differences, divide to find the average and then take the square root of the average.

If you square the differences, the final result is called the ***root mean squared error.*** If you take the absolute value of the differences, the result is called the ***mean absolute error.***

16. Calculate the root mean squared error or the mean absolute error for your line in Question 7. What does this value tell you?

You may recall the standard deviation or mean absolute deviation from earlier work. The root mean squared error is a measure of variability similar to standard deviation. The mean absolute error for a line is similar to the mean absolute deviation for a data set.

SUMMARY

- If there is a linear pattern in a set of data, you can summarize the relationship by drawing a line.
- You can determine how well a line fits a set of data by looking at the differences between the actual points and the predicted points on the graph.
- The difference between a predicted y-value and an actual y-value is called a ***residual.*** This difference is represented on the graph by the line segment that is the vertical distance from the point to the line.
- The overall error in predicted values can be measured by calculating the ***mean absolute error*** (the average of the absolute values of the differences or residuals) or the ***root mean squared error*** (the square root of the average of the squared differences or residuals).

Practice and Applications

17. In Florida, conservationists are concerned about the diminishing numbers of manatees, mammals similar to dolphins, that live in coastal waters. They are friendly and playful and seem to enjoy being near humans. During the late 1980s, the number of manatees whose dead bodies washed onto the shore increased dramatically. Examinations revealed that the manatees' bodies appeared to have been scraped by powerboat propellers. The Florida Department of Natural Resources investigated the number of powerboat registrations and found the following data.

Manatees and Powerboat Registrations in Florida

Year	Powerboat Registrations (thousands)	Manatees Found Dead on Shore
1977	447	13
1978	460	21
1979	481	24
1980	498	16
1981	513	24
1982	512	20
1983	526	15
1984	559	34
1985	585	33
1986	614	33
1987	645	39
1988	675	43
1989	711	50
1990	719	47

Source: data from *Decisions Through Data*, 1992

a. Describe any relationships or trends you can see from the data in the table.

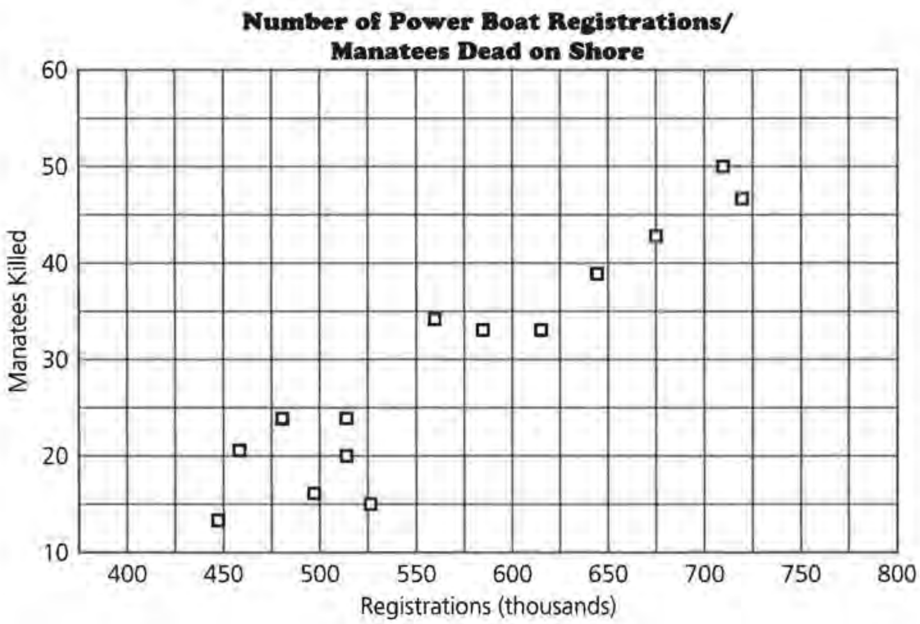
b. Make a scatter plot over time of the number of manatees found dead on shore. What does the scatter plot show?

c. Draw a line that will describe the deaths of the manatees as a function of time. Write the equation of your line. What does the slope of your line tell you?

18. The Florida conservationists observed that there seems to be a relationship between the number of powerboat registrations and the number of manatee deaths.

a. Using the data in the table, estimate the number of manatees found dead on shore as related to 550 thousand powerboat registrations.

b. Look at the following scatter plot to see if there really are any patterns as the conservationists claim. Compare the inferences you can make from the scatter plot over time in Question 17 with the inferences from the following scatter plot.



Source: data from *Decisions Through Data*, 1992

c. Use this scatter plot to estimate the number of manatees found dead on shore if there are 550 thousand powerboat registrations. How did you make your estimate?

19. Using *Activity Sheet 11*, draw on the scatter plot a line that you think summarizes the relationship between the number of manatee deaths and the number of powerboat registrations.

a. Write the equation of your line. What is the slope of your line, and what does it tell you about manatee deaths?

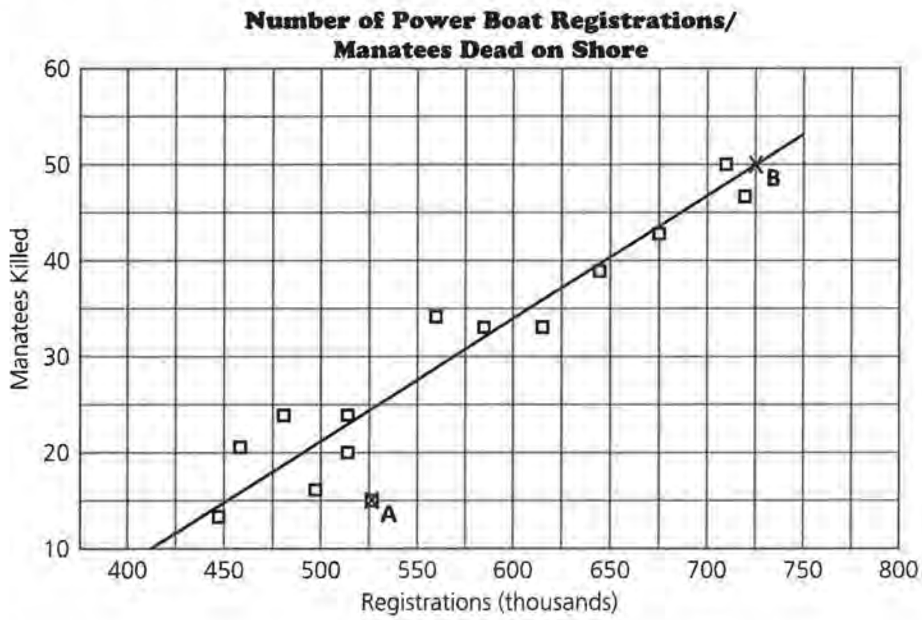
b. Using the equation of your line, predict the number of manatees found dead on the shore if there are 550 thousand powerboat registrations.

c. Which is the easiest to use to make a prediction: the table, the plot, the line, or the equation? Explain.

d. Find the y-intercept and the zero of your line. Use each in a sentence to describe the relationship between the number of powerboat registrations and the number of manatees found dead on shore.

20. Compare your line with the lines of others in your class. Who seems to have the *best* line? How did you decide?

21. Tua's line is shown on the following scatter plot.



a. What are the coordinates of A and B? Tell what each pair of coordinates represents.

b. If Tua uses her line to predict the number of manatees found dead on shore, would she predict too many or too few deaths for A? How can you tell?

c. When she made her predictions, Tua estimated her error by just counting the number of squares in the grid to represent the distance from a data point to her line. For example, she determined that A is about 2 squares away from the line so she concluded that her prediction was off by about 2 deaths. What would you tell Tua about her method?

22. Use the table, similar to the one below, on *Activity Sheet 11*.

a. Predict the number of manatees found dead on shore for the given number of powerboat registrations. Record your predictions in the table on *Activity Sheet 11*.

Powerboat Registrations (thousands)	Predicted Number of Manatees Found Dead on Shore
447	
460	
481	
498	
513	
512	
526	
559	
585	
614	
645	
675	
711	
719	

b. How did you make your predictions?

c. Who seems to have the best predictions in your group? How can you tell?

d. Calculate the root mean squared error or the mean absolute error for your predictions. What does the error indicate about your line?

e. Compare your error with the errors of your classmates. Who has the least error?

23. Write a letter to the Florida Department of Natural Resources outlining the problem involving powerboats and their negative effect on manatees as you have identified it. Provide statistical evidence to support your conclusion, and make some suggestions for resolving the issue.

24. Consider the perpendicular distance from a data point in a scatter plot to the line you have drawn to represent the relationship in the data. Do you think it would be useful to use this distance to measure a *best* line? Why or why not?

LESSON 8

Technology and Measures of Error

Finding residuals by hand is tedious and time consuming. How can you find them without so much work?

OBJECTIVE

Use technology to find a measure of error in predicting by using an equation.

Technology makes doing many calculations that are difficult or time consuming to do by hand easier. It is important, however, that you take time to understand the concepts involved.

INVESTIGATE

Using a Spreadsheet or Graphing Calculator

Refer to the data regarding the height and weight of seniors in Lesson 7, page 77. You can find the root mean squared error for the line $w = 170 + 7(h - 72)$, using either a spreadsheet software like Microsoft Works or a TI-83 graphing calculator.

If you are using *Microsoft Works*, open up the application to a new spreadsheet.

- In the first row, A1 through E1, enter the headings as follows:

A1	B1	C1	D1	E1
Height	Weight	Predicted Weight	Difference	Difference Squared

- Then enter the values for the height in column A (beginning in cell A2) and the corresponding values for weight in Column B (beginning with cell B2).
- In cell C2, enter your prediction equation, using A2 in place of h, with an equals sign (=) in front of the formula, =170+7*(A2–72).
- In cell D2, enter the formula =B2–C2.

- In cell E2, enter the formula **=(D2)^2.**

To label the results for the final computations:

- In D27, enter **Sum diff squared.**
- In D28, enter **Sum diff squared/*n*.** (*n* is the number of data points)
- In D29, enter **Root mean squared error.**

To enter the formulas for the final computations:

- In E27, enter **=Sum(E2: [location of last value in column E]).** In this example, there are 25 data points, so the entry in E27 is **=Sum(E2:E26).** (Recall that cell E1 does not contain a data point.)
- In cell E28, enter **=E27/25.**
- In cell E29, enter **=sqrt(E28).**

	A	B	C	D	E
1.	**Height**	**Weight**	**Predicted Weight**	**Difference**	**Difference Squared**
2.	72.5	195	=170+7*(A2–72)	=B2–C2	=(D2)^2
3.	70.5	137			
4.	72	156			
5.	70	145			
6.	71	175			
7.	70	165			
8.	73	195			
9.	76	155			
10.	67	143			
11.	76.5	155			
12.	76	192			
13.	69	149			
14.	71	225			
15.	66	132			
16.	63	97			
17.	61	145			
18.	61	108			
19.	64	106			
20.	66.5	125			
21.	64.5	140			
22.	69	120			
23.	68	130			
24.	61	105			
25.	69	140			
26.	62	110			
27.				Sum diff squared	=Sum(E2:E26)
28.				Sum diff squared/*n*	= E27/25
29.				Root mean squared error	=sqrt(E28)

The spreadsheet software will perform all of the operations for you, but you have to set it up. To make the formulas work in each of the columns C, D, and E, you must highlight each column individually. For example, highlight cell C2 and drag the highlight to the last data line. Under the **Edit** menu select **Fill Down.** This command will use the formula =170+7*(A2–72) in cell C2 to predict the weight for each height in cells A2 to A26 and store each result in cells C2 to C26, respectively. Repeat this process for columns D and E. The result in cell E29 will be the root mean squared error. Your final spreadsheet should look like the following one.

	A	B	C	D	E
1.	**Height**	**Weight**	**Predicted Weight**	**Difference**	**Difference Squared**
2.	72.5	195	173.5	21.5	462.25
3.	70.5	137	159.5	–22.5	506.25
4.	72	156	170	–14	196
5.	70	145	156	–11	121
6.	71	175	163	12	144
7.	70	165	156	9	81
8.	73	195	177	18	324
9.	76	155	198	–43	1,849
10.	67	143	135	8	64
11.	76.5	155	201.5	–46.5	2,162.25
12.	76	192	198	–6	36
13.	69	149	149	0	0
14.	71	225	163	62	3,844
15.	66	132	128	4	16
16.	63	97	107	–10	100
17.	61	145	93	52	2,704
18.	61	108	93	15	225
19.	64	106	114	–8	64
20.	66.5	125	131.5	–6.5	42.25
21.	64.5	140	117.5	22.5	506.25
22.	69	120	149	–29	841
23.	68	130	142	–12	144
24.	61	105	93	12	144
25.	69	140	149	–9	81
26.	62	110	100	10	100
27.				Sum diff squared	14,757.25
28.				Sum diff squared/*n*	590.29
29.				Root mean squared error	24.295884425

To find the mean absolute error, in cell E1 enter **Absolute Difference,** in E2 enter the formula, **=abs(D2),** in D27 enter **Sum absol diff,** and in D28 enter **Mean absol error.** You can delete the entries in D29 or in E29. The mean absolute error will appear in cell E28 when you use the same process you used to find the root mean squared error.

Discussion and Practice

1. Look at the preceding spreadsheet or the spreadsheet you generated.

a. What does the entry in cell C3 represent?

b. What does the entry in cell D9 represent?

c. Refer to the original scatter plot of the data in Lesson 7, Question 7, on page 78. What do the entries in columns C and D of the spreadsheet represent on the scatter plot?

2. Look at the spreadsheet on the previous page.

a. What is the root mean squared error for the equation $w = 170 + 7(h - 72)$?

b. Add the root mean squared error to $w = 170 + 7(h - 72)$. Then subtract the root mean squared error from the same equation. Use these two new equations to draw two lines on your scatter plot of the data (*Activity Sheet 9*).

c. What can the root mean squared error tell you about the predicted weight for a height of 70.5 inches?

When using a spreadsheet software, you can change the equation in your spreadsheet by entering a new slope and a new point into the equation in cell C2. Press **return,** then highlight C2 down to the last row in the C column in which you have data. Use the **Fill Down** command under **Edit** menu, and the recalculated results should appear on your spreadsheet. (Note: You may want to print your results each time to keep a record of your changes.)

3. Continue working with your spreadsheet.

a. Try a new equation in your spreadsheet to see how that affects the error.

b. Keep changing your equation to find as small a root mean squared error as you can.

c. What does the error you found in **b** tell you?

d. Why do you want as small an error as possible?

When using a TI-83 graphing calculator, press **STAT,** select **1:Edit,** press **ENTER.** Key in the heights in the **List 1** column and the weights in the **List 2** column. Press **Y=,** and at Y1=, key in **170+7(X–72)** as shown on the calculator screens at the top of the next page.

L1	L2	L3
72.5	195	------
70.5	137	
72	156	
70	145	
71	175	
70	165	
73	195	
L1 (1) = 72.5		

PLOT1 PLOT2 PLOT3
\Y1 = 170 + 7(x – 72)
\Y2 =
\Y3 =
\Y4 =
\Y5 =
\Y6 =
\Y7 =

The equation stored in Y1= will use the heights in List 1 for the X-values to generate the predicted weights. You can store the predicted weights in List 3 by following these steps.

- Press **STAT, 1,** to return to the Lists. Move the cursor to the top of the List 3 column. Press **ALPHA".**
- Press **VARS,Y-VARS, 1:Function, 1:Y1** (**2nd L1**) **ALPHA" ENTER.**

L1	L2	L3
72.5	195	------
70.5	137	
72	156	
70	145	
71	175	
70	165	
73	195	
L3 = "Y1(L1)"		

L1	L2	L3 •3
72.5	195	173.5
70.5	137	159.5
72	156	170
70	145	156
71	175	163
70	165	156
73	195	177
L3 = "Y1(L1)"		

The values in the List 3 column should be the weights predicted by using the equation stored in Y1=.

You can use List 4 in which to store the difference between each predicted weight and the actual weight by following these steps.

- Move the cursor to the top of the List 4 column. Press **ENTER.**
- At the bottom of the screen, at L4= (the location of the cursor) press **ALPHA" 2nd L2 – 2nd L3 ALPHA" ENTER** as shown on the following calculator screens.

L2	L3	L4
195	173.5	21.5
137	159.5	-22.5
156	170	-14
145	156	-11
175	163	12
165	156	9
195	177	18
L4 = "L2 – L3"		

To find the *absolute value* or the *squared difference,* define List 5 as either L5 = "abs L4" or L5 = "L4^2" as shown on the following calculator screens. The absolute value of the differences are shown in the left screen. The squared differences are shown in the right screen.

L3	L4	L5
173.5	21.5	21.5
159.5	-22.5	22.5
170	-14	14
156	-11	11
163	12	12
156	9	9
177	18	18
L5 (1) = 21.5		

L3	L4	L5
173.5	21.5	462.25
159.5	-22.5	506.25
170	-14	196
156	-11	121
163	12	144
156	9	81
177	18	324
L5 (1) = 462.25		

To find the mean absolute difference, press **2nd, QUIT.** Then press **2nd LIST MATH 3:mean.** At **mean** (, the location of the cursor, press **2nd L5**) (to specify the list you would like to use), and then press **ENTER.** The result will be the mean of the values in List 5 or the mean absolute difference as shown in the following calculator screens.

```
mean(L5)
                    18.54
```

If you choose to use the squared difference, press **2nd** √ **2nd LIST MATH.** Select **3:mean(.** Press **ENTER.** Press **2nd L5)** **ENTER.** The result will be the mean of the squared differences in List 5, or the root mean squared error as shown on the following calculator screens.

```
NAMES OPS MATH
1: min(
2: max(
3: mean(
4: median(
5: sum(
6: prod(
7↓stdDev(
```

```
mean(L5)
                 24.29588443
```

4. Refer to the calculator screens shown below.

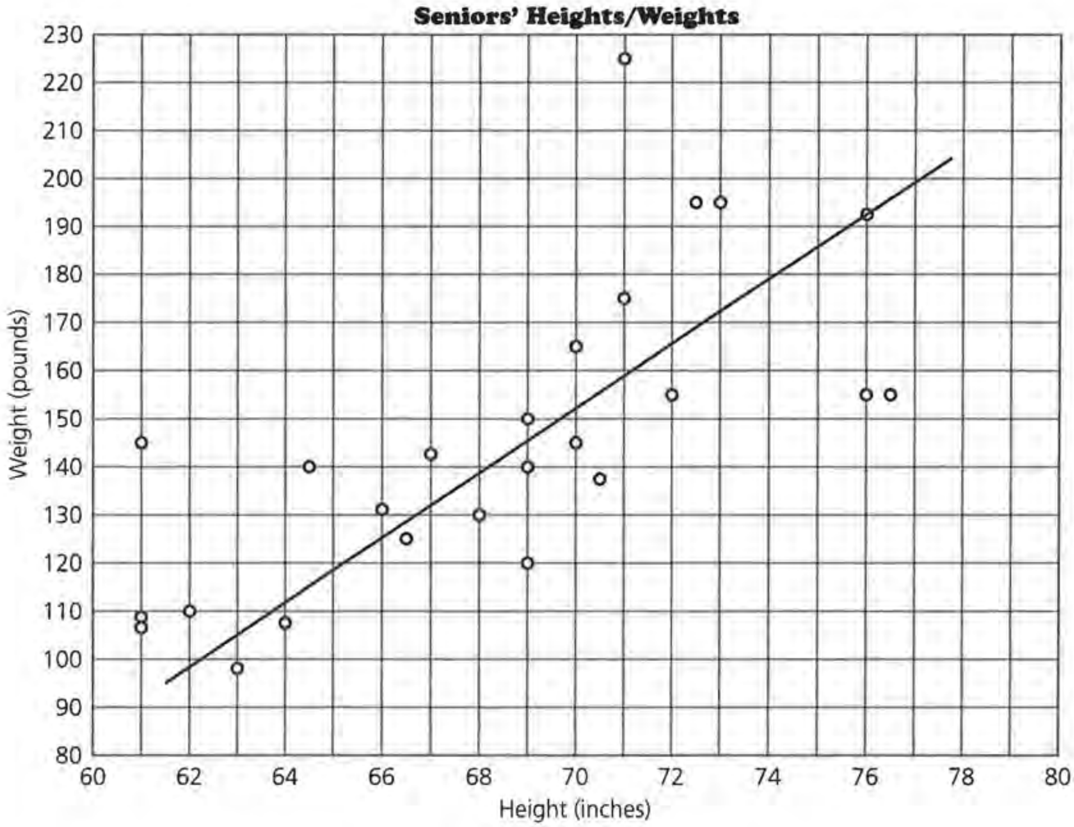
a. What does the entry in L3(4) represent?

b. What does the entry in L4(3) represent?

L3	L4	L5
173.5	21.5	21.5
159.5	22.5	22.5
170	-14	14
156	-11	11
163	12	12
156	9	9
177	18	18
L3 (4) = 156		

L3	L4	L5
173.5	21.5	21.5
159.5	22.5	22.5
170	-14	14
156	-11	11
163	12	12
156	9	9
177	18	18
L4 (3) = -14		

c. Refer to the original scatter plot of the data, repeated on the next page. What do the entries in Lists L3 and L4 represent on the scatter plot?



d. What is the root mean squared error for the equation $w = 170 + 7(h - 72)$? What does this tell you about your predicted weight for a height of 70.5 inches?

5. To find the error for a new equation, enter the new equation in Y1. Look at the graph. The new predicted values will be in L3. The difference between the observed and predicted values will be in L4. Press **2nd ENTER** and **ENTER** again to replay the mean error.

a. Try a new equation, and see how that affects the error.

b. Keep changing your equation to find a smaller error.

c. What does the error you found in **b** tell you?

d. Why do you want the smallest error possible?

SUMMARY

Spreadsheet software or a graphing calculator can be used to quickly find a measure of error for a prediction equation.

- Once you have the equation of the line, you can generate predicted y-values.
- The root mean squared error will give you an "average" interval above and below a predicted y-value.

- With spreadsheet software, you can change one of the values in the equation and fill down. The resulting new error appears automatically. This feature lets you easily experiment with different equations to find one that reduces the error.
- With a TI-83 graphing calculator, you have to change the equation in Y1= each time and recalculate the error from the lists generated by that equation.

Practice and Applications

6. The following table shows the recommended weight for females and males as a function of height for ages 19 through 34.

Recommended Weight for Females and Males, Ages 19–34

Height (feet and inches)	Weight (pounds) Female	Weight (pounds) Male
5′0″	97	128
5′1″	101	132
5′2″	104	137
5′3″	107	141
5′4″	111	146
5′5″	114	150
5′6″	118	155
5′7″	121	160
5′8″	125	164
5′9″	129	169
5′10″	132	174
5′11″	136	179
6′0″	140	184
6′1″	144	189
6′2″	148	195
6′3″	152	200
6′4″	156	205
6′5″	160	211
6′6″	164	216

Source: data from *Nutrition and Your Health: Dietary Guidelines for Americans*

a. Plot the suggested weight for males as a function of height (*height, weight*), and determine whether a line would be appropriate to describe the relationship. If so, draw the line and find its equation.

b. Use spreadsheet software or a graphing calculator to find the root mean squared error for your equation. What does the root mean squared error indicate about any prediction you make using your equation?

c. If someone is 4 feet 8 inches tall, what suggested weight would you expect to appear in the *Dietary Guidelines?*

d. Add the root mean squared error to your equation, and graph the new equation on your scatter plot. Subtract the root mean squared error from the original equation, and graph that equation on your scatter plot. How many points lie outside this band? What does this indicate?

7. What is the relationship between calories and the amount of fat in foods from fast-food restaurants? The following table shows the number of calories and grams of fat in food items offered at several fast-food restaurants.

Fast-Food Restaurant Items

Restaurant	Food Item	Fat (grams)	Calories
McDonald's	McLean Deluxe without cheese	10	320
	Quarter Pounder with cheese	28	510
	Chicken Salad	4.5	162
Burger King	BK Broiler (no sauce)	8	267
	Chunky Chicken Salad	5	172
	Whopper with cheese	44	706
Pizza Hut	Cheese pizza (two slices)	18	492
	Pan Supreme Pizza	30	589
Taco Bell	Bean Burrito (no cheese)	14	447
	Chicken Burrito (no cheese)	12	334
	Taco Salad (with shell)	61	905
	Nachos-Bell Grande	35	649

a. Plot the calories as a function of the grams of fat (*fat, calories*), and draw a line that seems to represent the relationship. Find the equation of your line.

b. Find the y-intercept and zero of your line, and indicate how to interpret each in the context of fat and calories.

c. Use spreadsheet software or a graphing calculator to find the mean absolute error.

d. Compare your line with the lines of your classmates, and determine who has the *best* line.

e. Write a paragraph summarizing what your equation means and what the error indicates about using your equation to predict values.

8. Describe how you can determine a measure of error for a prediction equation. Give an example, and explain how it works.

LESSON 9

Lines, Lines, and More Lines

Do you have to draw several lines and test each one for error?

How can you draw a fairly good line to begin with?

If you look at some of the possible lines summarizing the relationship of the data, how can you be sure that the lines are fairly similar?

There are many possible lines that can be used to describe a set of data. There are some common techniques that are used to draw different lines. One typical line for a set of data is called a *median fit line*. The technique used to draw the median fit line is based on ordering the data points and counting to find median points that represent the data in different sections of the scatter plot.

OBJECTIVES

Summarize data by drawing a median fit line. Solve a linear equation.

INVESTIGATE

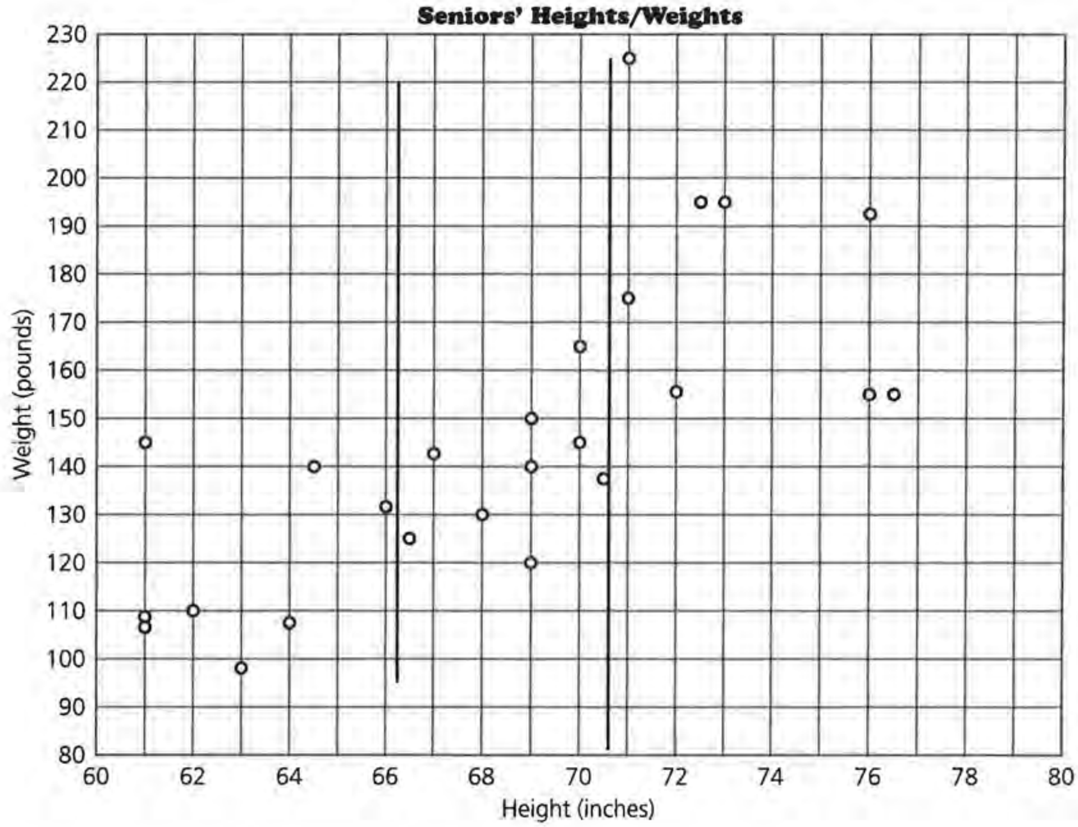
Seniors' Heights and Weights

Consider the seniors' height/weight data once more as reproduced on the following page.

Seniors' Heights/Weights

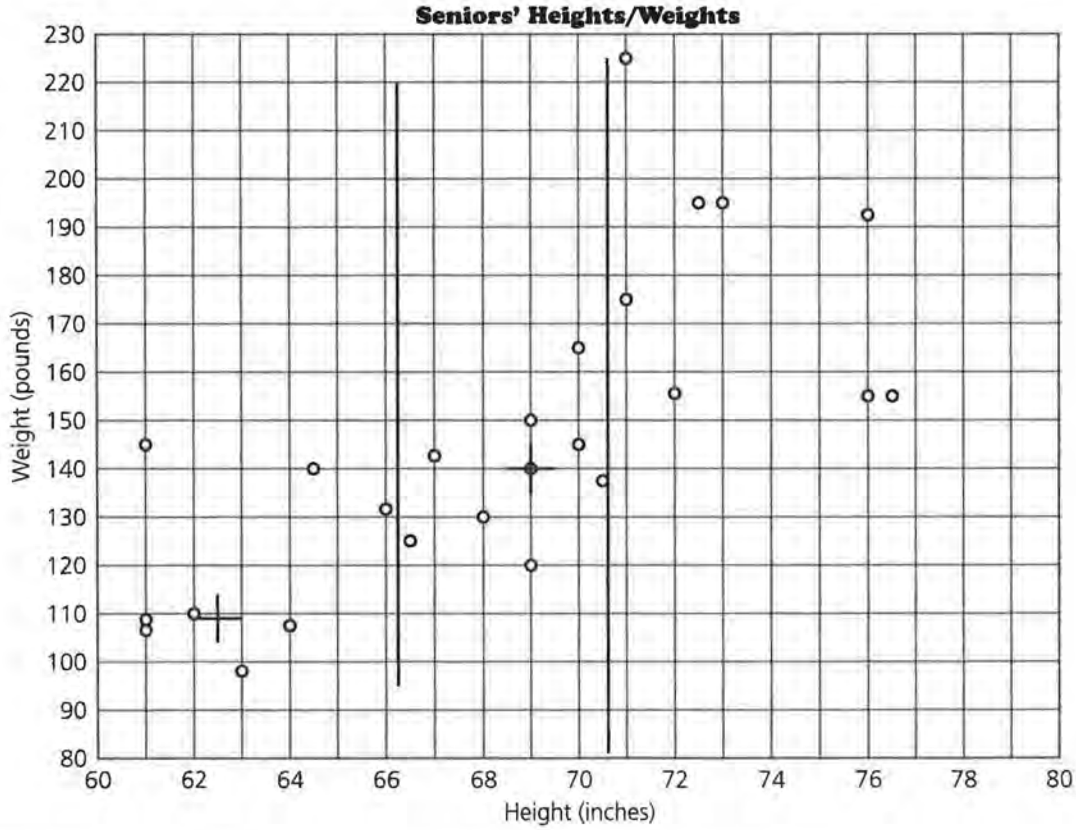
Males' Height (inches)	Weight (pounds)	Females' Height (inches)	Weight (pounds)
72.5	195	66	132
70.5	137	63	97
72	156	61	145
70	145	61	108
71	175	64	106
70	165	66.5	125
73	195	64.5	140
76	155	69	120
67	143	68	130
76.5	155	61	105
76	192	69	140
69	149	62	110
71	225		

Recall that you only need two points to draw a line, but sometimes it is important to use a third point as a "checkpoint." To help you find these three points, draw vertical lines perpendicular to the horizontal axis separating the data points into three regions of the scatter plot. (The vertical lines you draw should not contain any data points.) Each of the three regions should contain, as nearly as possible, an equal number of points. In this example there are 25 points, as shown in the following scatter plot, and because $25 \div 3 = 8$ with a remainder of 1, put 8 points in the first region, 9 in the middle, and 8 in the third.



You can determine a representative, or median, point for each region by finding a median value for the height, the x-value, and a median value for the weight, the y-value.

One way to find these median points is to look at the scatter plot. For this example, in the first region (to the left), the median height will be halfway between the fourth and fifth horizontal coordinates, or at 62.5. (Notice that there are 4 points to the left of 62.5 and 4 points to the right of 62.5.) The median weight will be halfway between the fourth and fifth vertical coordinates, or at 109. (Notice that there are 4 points above 109 and 4 points below 109.) Thus, the median point for the first region has coordinates (62.5, 109), as shown in the next scatter plot. (Notice that in this example, for the first region, the median point is *not* one of the data points.)



Another way to find these median points is to make an ordered list of the coordinates of the points in each region and find the median x- and y-values as shown in the table below.

First Region (*height, weight*)	**Heights (in order)**	**Weights (in order)**
(61, 145)	61	97
(61, 108)	61	105
(61, 105)	61	106
(62, 110)	62	108
	Median = 62.5	**Median = 109**
(63, 97)	63	110
(64, 106)	64	132
(64.5, 140)	64.5	140
(66, 132)	66	145

In the second or middle region, the median height is an actual horizontal coordinate because there is an odd number (9) of horizontal coordinates for the points in this region. The fifth horizontal coordinate is at 69. (Having several points with horizontal coordinates of 69, does not change your result.) The fifth vertical coordinate is 140; therefore the coordinates of the representative or median point for the second region is (69, 140).

The median point for the third, or right, region is found in the same way as the point for the first, or left, region.

Discussion and Practice

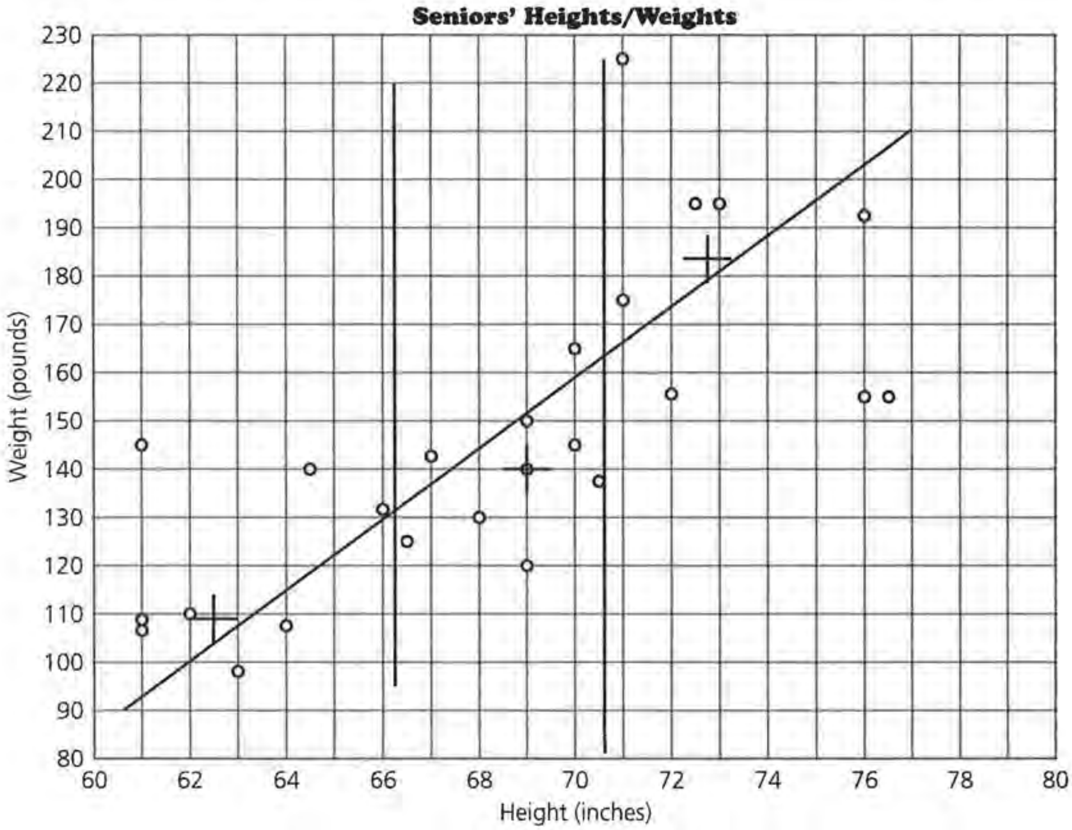
1. Refer to the table of *Seniors' Heights and Weights.*

  a. Make an ordered list of the coordinates of the data points in the middle region, and verify that the median height is 69 inches and the median weight is 140 pounds.

  b. What are the coordinates of the median point in the right region?

  c. How would you separate the data points into three regions if there are 26 points altogether?

You can use the median points in the left and right regions of the scatter plot to determine the slope of the median fit line. Place your ruler on the two median points, and while keeping it parallel to the line between them, slide your ruler $\frac{1}{3}$ of the vertical distance toward the median point in the middle region. Now draw this line on your scatter plot, as shown on the next page. (Even though you are approximating $\frac{1}{3}$ of the distance, your line will still be fairly accurate.)



The line you have drawn is called a *median fit line,* or a *median-median line,* and can be used to summarize the relationship between the heights and weights of the 25 seniors. To find the equation of the median fit line, use any two points on the line.

2. Consider the preceding scatter plot.

a. Find an equation for the median fit line shown in the preceding scatter plot.

b. What is the rate of change, or slope of the line, and what does it tell you?

c. Use the equation of your line to predict the weight of a person who is 74 inches (6 feet, 2 inches) tall.

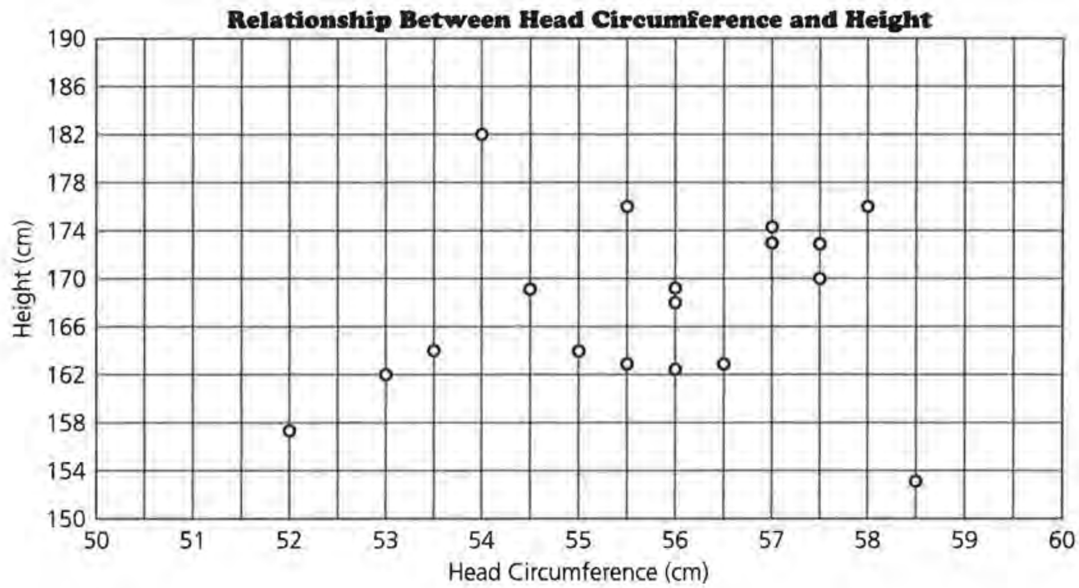
d. What is the y-intercept of the line? What does the y-intercept mean in terms of the data? Does this make sense?

3. Refer to the median fit line of the preceding scatter plot.

a. Earlier you found the root mean squared error or the mean absolute error for a line. Find either of these values for the median fit line. How does the error for the median fit line compare with the error for the line you drew by eye at the beginning of Lesson 8.

b. Find the weight of a person who is 68 inches tall. What does the error indicate about your prediction?

4. Following is a scatter plot showing (in centimeters) both the circumference of the head and the height of some high-school freshmen.



a. Identify any of the points that look as though they might be outliers. Why do you think these points might be outliers?

b. The class found that two of the class members had converted inches to centimeters incorrectly when they were measuring their height. The correct points should be (58.5, 172) and (54, 164). Using *Activity Sheet 12*, adjust the scatter plot accordingly.

c. The measurements for the rest of the class are shown in the table below. Add these points also to the scatter plot on *Activity Sheet 12*.

Freshmen's Head/Height Measurements

Head Circum. (centimeters)	Height (centimeters)	Head Circum. (centimeters)	Height (centimeters)
54	160	54.5	164
55	166	53	156
59	178	53.5	158.5

d. Draw a median fit line on *Activity Sheet 12* to summarize the data in the corrected scatter plot. Write the equation of your median fit line. What is the slope, and what does it tell you in terms of the data?

e. Use your median fit line to predict the height of a person with a head measure of 55.5 centimeters. How confident are you about your prediction? Describe what adjustments you could make to improve your confidence in the accuracy of your prediction.

5. Look at the scatter plot in Question 4 showing the two incorrect data points. The scatter plot is also reproduced on *Activity Sheet 13*.

a. Plot the six extra points on this scatter plot and draw a median fit line. Write the equation for this median fit line. How does this equation compare with the one you found in Question 4d?

b. Comment on this statement: A median fit line will not be changed much by points that are outliers.

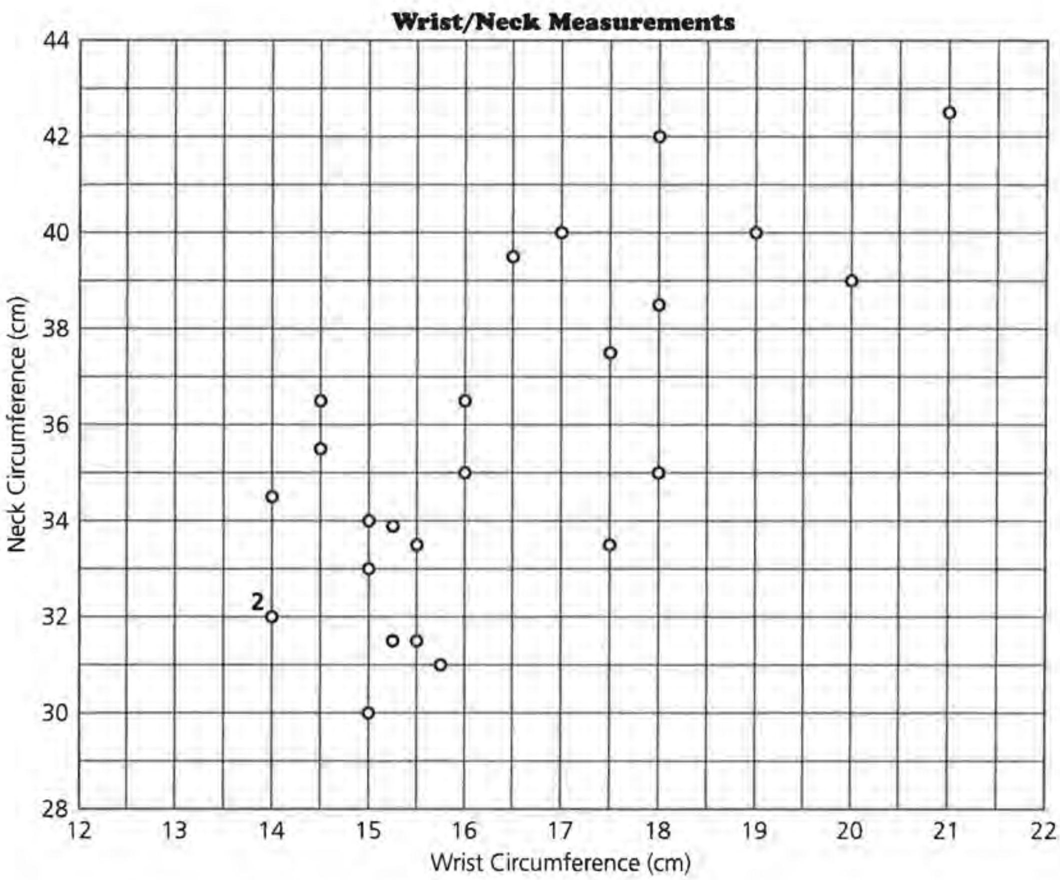
6. Is there any relationship between the size of a person's wrist and the size of that person's neck? A class of students measured their wrists and necks in centimeters as shown in the following table.

Wrist/Neck Measurements

Wrist (centimeters)	Neck (centimeters)	Wrist (centimeters)	Neck (centimeters)
15	30	17.5	33.5
15.7	31	15	33
14	34.5	16	35
15	34	14	32
15.5	31	17	40
18	35	16.5	39.5
15.3	31.5	17.5	37.5
14	34.5	15.3	33.8
19	40	18	42
14.5	35.5	21	42.5
16	36.5	20	39
18	38.5	14.5	36.5
15.5	33.5		

a. Why do you think the measurements are in centimeters and not in a larger unit of measure?

b. The data are shown on the following scatter plot and also on *Activity Sheet 14*. Using the activity sheet, draw a median fit line on the scatter plot of the data. Write the equation of the median fit line. Describe the rate of change, and indicate what this tells you in the context of the data.



c. Predict the neck size of a person who has a 17-centimeter wrist measurement.

d. Find either the mean absolute error or root mean squared error for the median fit line you drew. What does this tell you about your prediction? How well do you think the line summarizes the data?

Sometimes you want to know what wrist size would predict a given neck size. To find out, you can use the graph of your line on the scatter plot, a table of values, or the equation of your line.

7. Suppose the equation relating the wrist/neck data is $N = 33 + 1.5(W - 14.5)$. What wrist size would predict a neck size of 35 centimeters?

   a. Explain how you can use a table of values to find the solution.

   b. Explain how the graph will help you find the solution.

Suppose you use algebraic properties to find the wrist size that would predict a neck size of 35 centimeters. You can rewrite the equation in an equivalent form and solve.

If the predicted neck size is 35, let $N = 35$ in the following equation:

$$N = 33 + 1.5(W - 14.5)$$

$$35 = 33 + 1.5(W - 14.5)$$

Distribute 1.5 and combine like terms.

$$\begin{aligned} 35 &= 33 + 1.5(W - 14.5) \\ &= 33 + 1.5W - 1.5(14.5) \\ &= 33 + 1.5W - 21.75 \\ &= 11.25 + 1.5W \end{aligned}$$

Since $35 = 11.25 + 1.5W$, you must find a quantity that when added to 11.25 is 35.

$$35 = 11.25 + ?$$

Thus 23.75 has to equal $1.5W$ because

$$23.75 = 11.5 + 23.75$$

So: $1.5W = 23.75$

$$W = 23.75 \div 1.5$$

$$W = 15.83$$

Therefore, a person with a wrist size of about 16 centimeters would have a neck size of 35 centimeters.

8. Using the equation in Question 7, what wrist size would predict a neck size of 37 centimeters?

SUMMARY

The median fit line for data points is one way to find a *good* line to represent the data.

- You can find the median fit line by separating the data points into three regions and finding the median point that represents the data points in each region.
- The median fit line is not sensitive to outliers because the median of a data set is not affected by extreme values.
- The median fit line can be used to predict y from x, to interpolate or read between the known data points, to extrapolate or read beyond the known data points (recognizing the assumptions made in using the value), or to act as a summary equation for the data.
- You can predict outcomes by using the graph of the median fit line, by inspecting sets of ordered pairs generated by the equation of the median fit line, or by substituting a known value into the equation and solving for the unknown value.

Practice and Applications

9. Measure the arm span (in inches) and the height of each member of your class. (Arm span is the distance from the fingertip of your left hand to the fingertip of your right hand when your arms are outstretched parallel to the floor.)

a. Plot the data (*height, arm span*), and draw a line. Find an equation to describe the relationship.

b. Use your equation to predict the arm span of someone who is 63 inches tall.

10. Suppose $y = 172x - 32$. Find each of the following.

a. x, when y is 14 **b.** y, when x is 22

c. x, when y is 0 **d.** y, when x is 0

11. Find the value for x in each equation.

a. $15 = 4 + 3\ (x - 8)$ **b.** $8 = 14 + {}^{-}3(x - 2)$

c. $750 = 182 + 7x$ **d.** $330 = {}^{-}120 - 7x$

12. The following table shows the mean rainfall per year and the mean percent of sunshine in selected cities around the United States.

Sun/Rain in U. S. Cities

City	Mean Rainfall (inches)	Mean % Sun	City	Mean Rainfall (inches)	Mean % Sun
Los Angeles, CA	14	73	New Orleans, LA	57	59
Salt Lake City, UT	15	70	Nashville, TN	46	57
Phoenix, AZ	7	86	Jackson, MS	49	60
Las Vegas, NV	9	84	Mobile, AL	60	67
San Francisco, CA	20	67	Atlanta, GA	61	48
Denver, CO	16	70	Charlotte, NC	66	43
Wichita, KS	31	65	Raleigh, NC	60	43
Oklahoma City, OK	31	67	Miami, FL	66	60
Albuquerque, NM	8	77	St. Louis, MO	58	36
Houston, TX	48	57	Louisville, KY	57	43
Little Rock, AR	49	63	Norfolk, VA	63	45

Source: data from *The Greening of America*, 1970

a. Plot (*rainfall*, *sunshine*) and find the median fit line. Write its equation.

b. What do the slope, y-intercept, and zero represent in terms of the data?

c. What amount of rain would you expect for an average of 75% sunshine?

d. How well do you think the line represents the data? Explain your answer.

13. For the following sets of data, plot the points and find a median fit line for each. Write a paragraph, describing the relationship between the variables for each set of data. Include in your paragraph a discussion of the slope, y-intercept, and zero of the median fit line.

a. The prices of used Ford Mustangs are shown in the following table.

Prices of Used Ford Mustangs from Classified Ads in 1993

Year	Age	Price
1980	14	$900
1980	14	575
1983	11	1,950
1985	9	2,695
1986	8	1,795
1986	8	1,695
1986	8	2,500
1987	7	7,495
1988	6	4,300
1988	6	6,800
1988	6	8,495
1990	4	7,295
1990	4	11,995
1993	1	8,493
1994	1	12,799

b. The following table, which has been used by the Federal Bureau of Investigation in shoe print studies, shows the heights and the shoe sizes of males.

Male Height and Shoe Size

Height (inches)	Shoe Size	Height (inches)	Shoe Size
65	6	70	11
66	6.5	72	11
68	7	72	10.5
67	7.5	73	11.5
68	8	73	12
68	8.5	75	12.5
69	9	77	13
70	9.5	79	14

Source: *Footwear Impression Evidence*, 1990.

14. Collect a data set with at least 15 pairs of data that you think may have a linear relationship, and then plot the data. Using what you learned in this unit, answer each of the following questions.

a. Draw a line to summarize the relationship, and explain how it can be used to predict values.

b. Do you think your hypothesis about a linear relationship is correct? Why or why not?

c. How reliable do you think a prediction based on your line will be? Explain your answer.

d. Find an equation for the line you drew. Write two questions that could be answered by using your equation.

A Measure of Association

"Grades are correlated with homework!" "The use of Vitamin E is correlated with a decrease in strokes." "There is a correlation between exercise and human longevity."

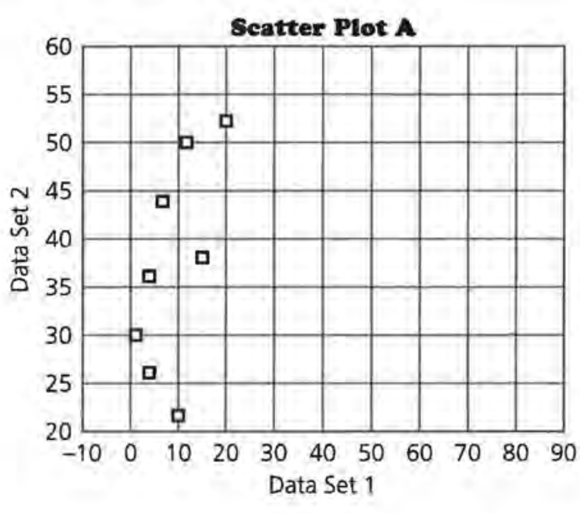
What do these statements mean?

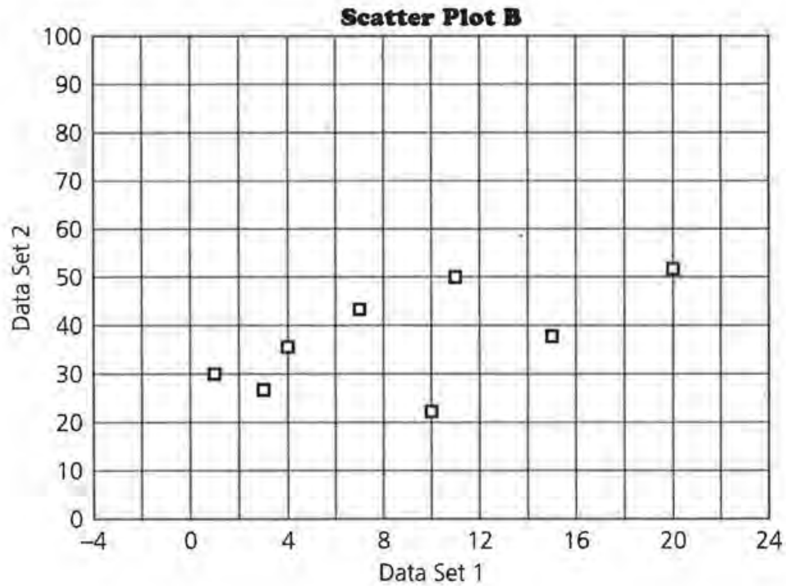
What does it mean when two variables are correlated?

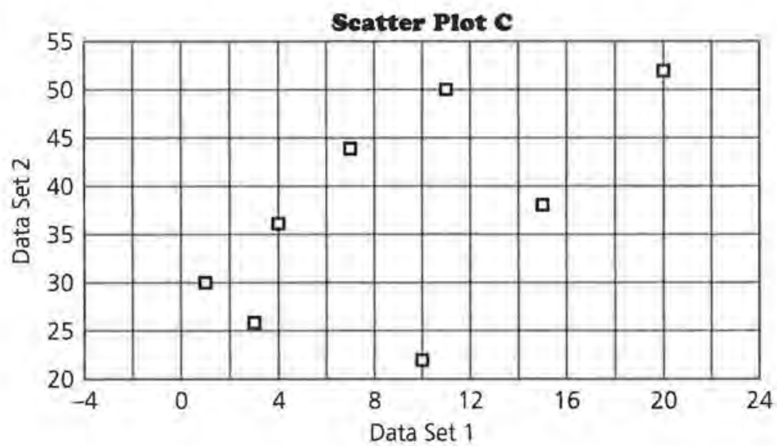
The scales on the axes can distort your perception of the pattern in a scatter plot. Scatter Plots A, B, and C show the same data graphed on grids with different scales. Which of the scatter plots looks most linear? Although it is important to look at the scatter plot to see if the data seem linear, a numerical measure can help you make a judgment about the relationship between the variables that is not dependent on the scales used in the scatter plot.

OBJECTIVES

Determine the strength of a linear relationship by finding a measure of correlation. Understand the connection between correlation and cause/effect.







Correlation measures the strength of a linear relationship. If the data points form a pattern that looks very much like a straight line, the correlation is called *strong*. If the data points are scattered or form a circle or large oval, the correlation is called *weak*.

What is the strength of the association or correlation between height and weight? Should you use a line to predict the weight of a person if given a height?

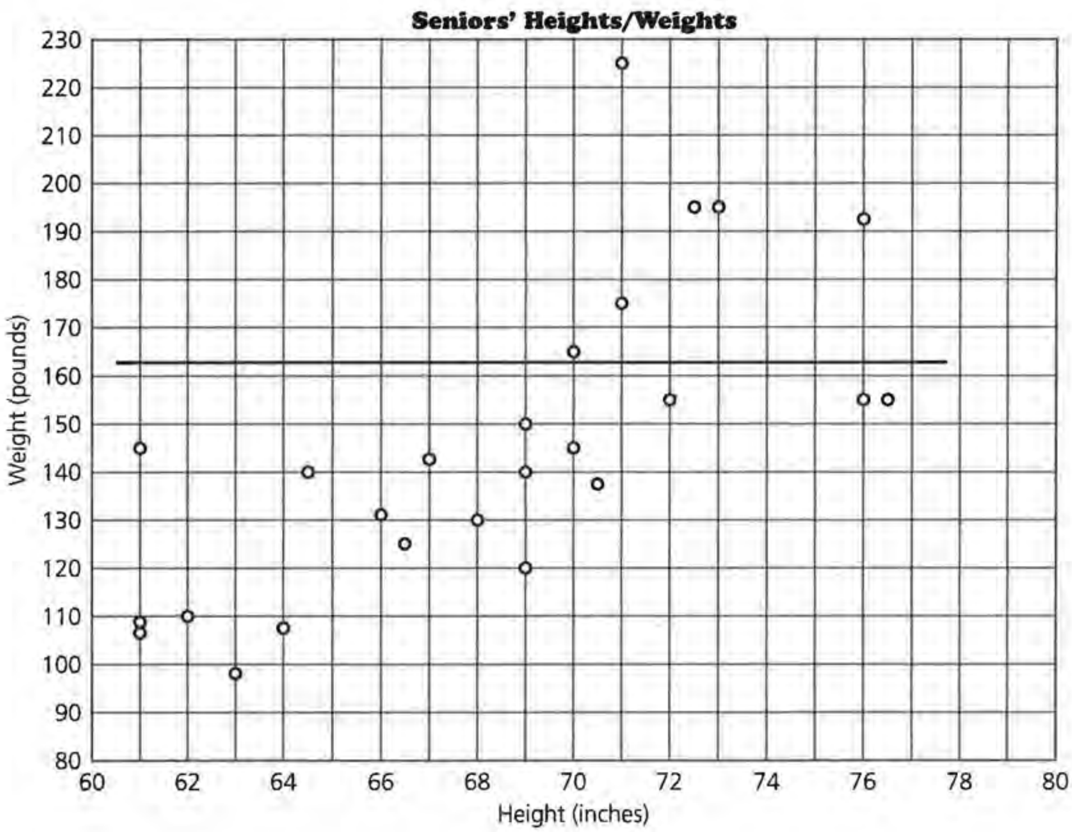
INVESTIGATE

Correlation Between Height and Weight

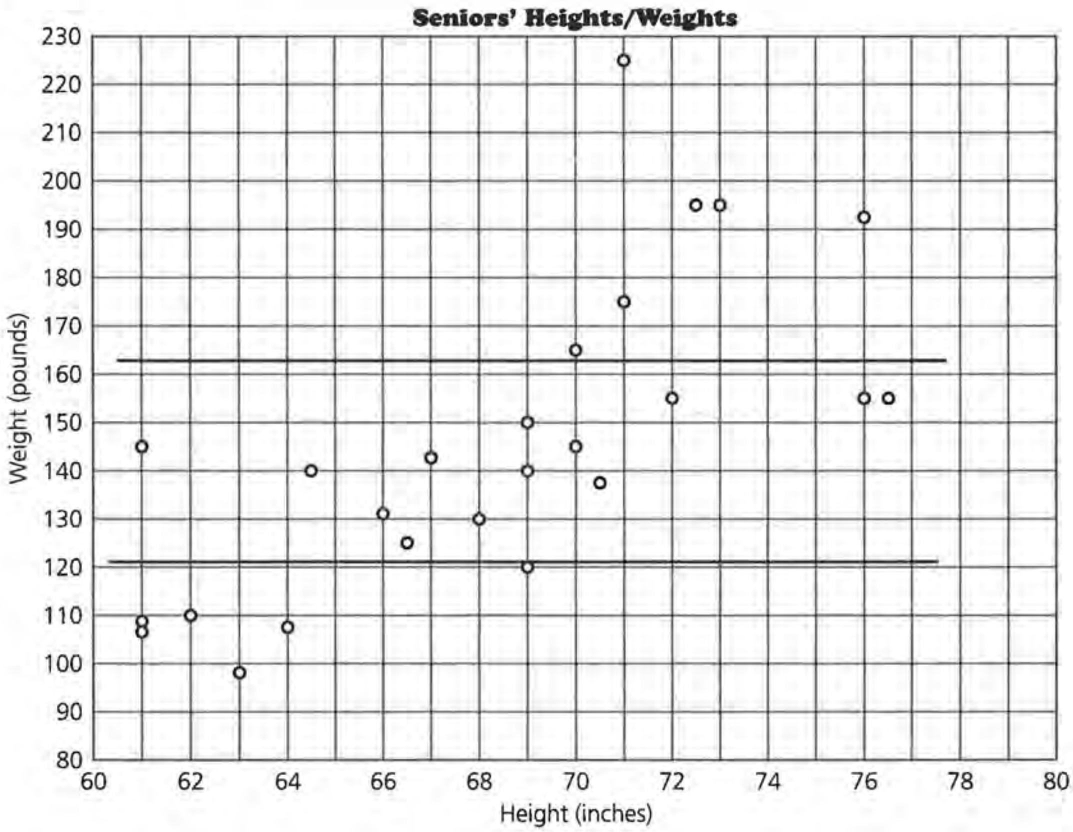
In Lesson 9, you fit a line to the data by separating a scatter plot into three regions. To estimate the value of the correlation coefficient, you also begin by separating the scatter plot into regions. This time, however, you will separate the scatter plot into nine regions, like a tick-tack-toe grid.

First, count the number of data points (N) and divide N by 4 (because you are interested in the 4 corner regions). Recall that there were 25 people in the (*height, weight*) scatter plot, so in this example, $25 \div 4 = 6.25$. Disregard the decimal part (0.25) of the number and use the integer part, called I. In this example $I = 6$. (This function I is called the ***greatest integer function*** and can be written as $[6.25] = 6$.)

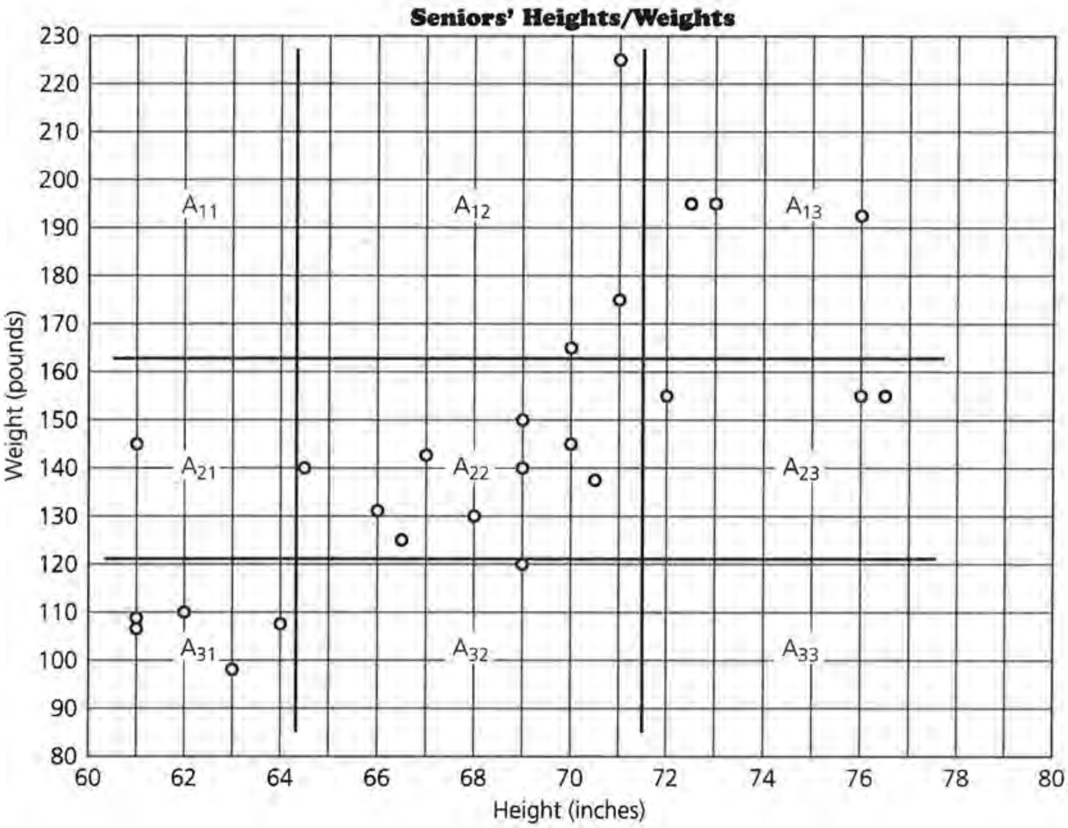
Count 6 or fewer points from the top of the scatter plot and draw a horizontal line as shown here. (If there are "ties," or two points with the same y-coordinate, count fewer than the value of I.)



Now count 6 or fewer points from the bottom of the scatter plot and draw another horizontal line as shown in the following scatter plot.



Count 6 or fewer points from the left and 6 or fewer points from the right, and draw two vertical lines. Any line you draw should not pass through a data point. Now you have a tick-tack-toe grid on your scatter plot as shown on the next page.



You can label cells in your tick-tack-toe grid, or entries in your matrix, by indicating the positions using rows and columns. (Rows go across and columns go up and down.) Cell A_{21} on your scatter plot indicates the second row and the first column on the grid. In the matrix, your entry for A_{21} is 1 because there is one data point in the A_{21} cell.

A_{11}	A_{12}	A_{13}	For your example,	0	3	3
A_{21}	A_{22}	A_{23}	the number of points	1	9	3
A_{31}	A_{32}	A_{33}	in each cell is	5	1	0

Note that A_{11} is 0 because there are no data points in the A_{11} cell of your tick-tack-toe grid.

r is used to represent the correlation. To estimate the correlation *r*, use this formula:

$$r = \frac{(A_{13} + A_{31}) - (A_{11} + A_{33})}{2I}$$

So, $r = \frac{8-0}{2(6)}$

$= 8 \div 12 = 0.6...$, or about 0.67

Suppose you have to interpret a correlation of 0.67. You might first think about some extreme cases.

Discussion and Practice

1. Suppose you are given 20 points that form a straight line with a positive slope.

a. Make a sketch of what the scatter plot might look like. What is *I*?

b. How many points will lie in A_{13}? In A_{11}? In A_{31}? In A_{33}?

c. Find *r*.

d. Will *r* change if you are given 32 points to determine the line instead of 20?

2. Jo-Jo plotted points for three straight lines and found the correlation for each. The conclusions he drew are shown here. Decide whether or not he was correct in each case, and explain why you agree or disagree with him.

a. If the points are in a straight line with a positive slope, then $r = 1.00$.

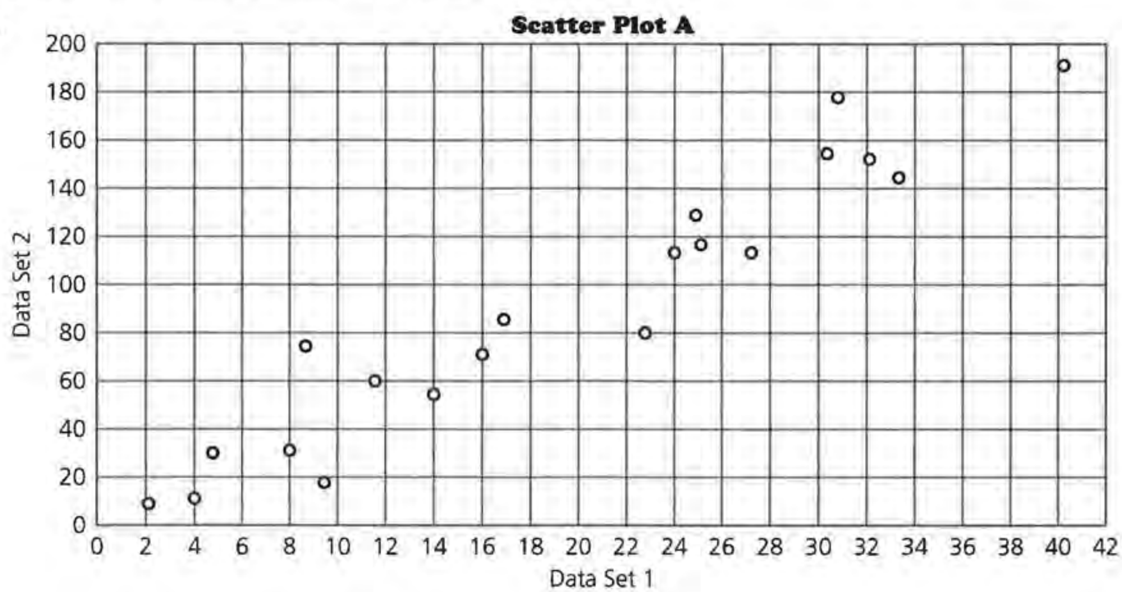
b. If there are no points in cells A_{11} and A_{33}, then $r = 1.00$.

c. If the points are in a straight line with a negative slope, then $r = {}^{-}1.00$.

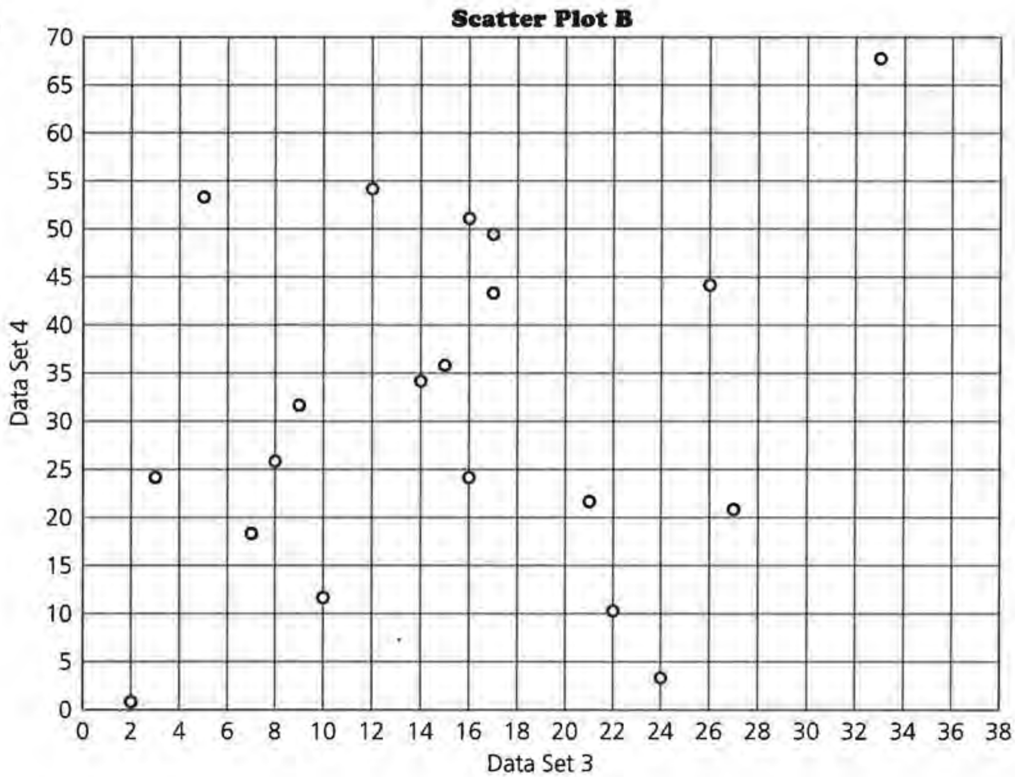
d. If the total number of points in A_{11} and A_{33} is the same as the total number of points in A_{13} and A_{31}, the correlation is 0.1.

3. Estimate the correlation for each of the following two scatter plots. Use *Activity Sheet 15* for your work.

a.



b.



4. Consider a set of points with a negative correlation.

a. If a set of points has a negative correlation, describe a possible line of fit for those points. How would a negative correlation affect the numbers in the cells of the matrix?

b. Give a set of points with a negative correlation and estimate *r*.

c. Describe the difference between positive correlation and negative correlation.

- A correlation close to 1 indicates a strong positive relationship between the variables. The data points show a pattern close to a straight line with a positive slope.
- A correlation close to $^-1$ indicates a strong negative relationship between the variables. The data points show a pattern close to a straight line with a negative slope.
- If the correlation is close to 0, there is no relationship between the variables. The data points show a circular pattern or a pattern that is not linear.

5. Refer to the table below.

a. Find at least 18 sets of possible values for the length, width, and area of a rectangle with a perimeter of 50 feet. Then add your values to the table started below.

Rectangle with P = 50 feet

Length (feet)	Width (feet)	Area (square feet)
5	20	100

b. Explain why 5 and 20 are possible values.

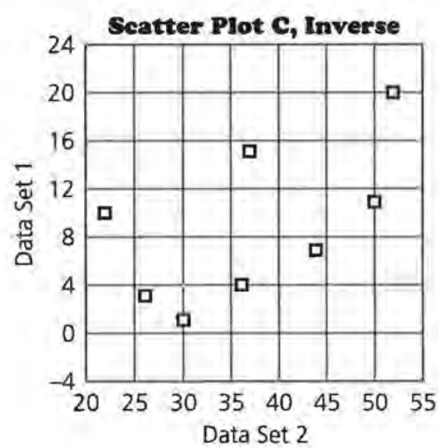
c. Plot each of the following sets of ordered pairs: (*length, width*) and (*length, area*). Then estimate the correlation. If the data appear linear with a correlation close to 1, draw a line that fits the data and write its equation.

d. What conclusion can you make about correlation based on your findings in c?

6. Refer to the scatter plots at the beginning of this lesson. The approximate correlation for Scatter Plot A is 0.25.

a. What do you think the correlation will be for Scatter Plots B and C? Explain your answer.

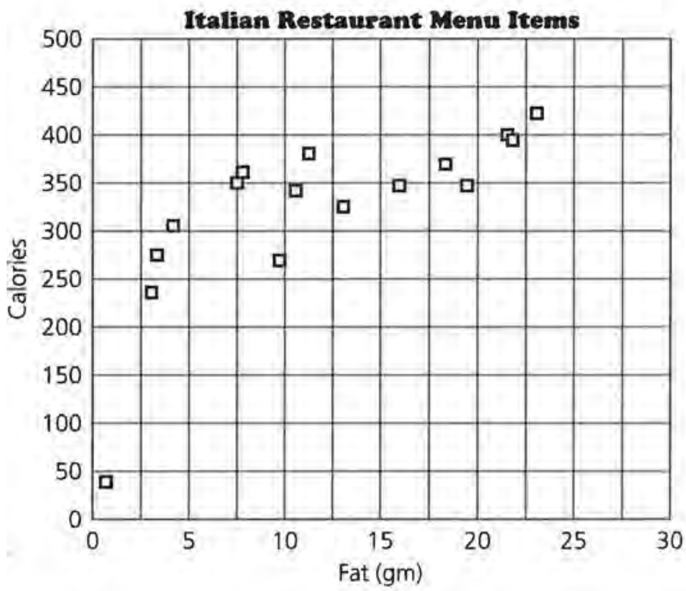
b. The following scatter plot is the *inverse* of Scatter Plot C. The inverse is formed by reversing the x- and y-values in each of the ordered pairs so that (10, 22), for example, has been plotted (22, 10). What is the estimated correlation coefficient for the data in this inverse scatter plot?



7. The following table shows the grams of fat and the number of calories for several entrees at an Italian restaurant. These data points are graphed on the scatter plot that follows the table.

That's Amoré Restaurant

Menu Item	Fat (grams)	Calories	Menu Item	Fat (grams)	Calories
Linguini Primavera	2.9	235	Chicken Piccata	16.1	348
Linguini with Scallops	3.1	276	Italian Frittata	23	424
Linguini with Shrimp	4.1	304	Sicilian Pork	11.2	379
Classic BLT	19.6	347	Florentine Chicken	7.9	362
TLT (Turkey)	13.1	328	Veal Piccata	21.9	394
Chicken Primavera	7.5	350	Sweet and Sour Chicken	10.4	341
Eggplant Parmigiana	9.7	269	Veal Parmigiana	21	450
Poached Roughy	18.4	370	Breadsticks	0.6	40



a. Estimate the correlation.

b. Describe the relationship between the grams of fat and the number of calories in the menu items shown in the scatter plot.

c. Find the median fit line and write the equation. What does the slope of the median fit line indicate about the data?

d. If you know the correlation coefficient, what can you say about the slope of the line for the data?

Correlation is often confused with *cause and effect.* A strong correlation between two variables does not necessarily mean that in the relationship one variable caused the other one to change. The changes in the variables could have been caused by some other factor or the correlation could be coincidence.

8. There is a strong correlation in each of the following. Discuss what you think might cause the association, and discuss whether the statement in **c** makes sense. Is there a third factor in statements **c** and **d** that could have caused both variables to change?

a. Small children have small feet and cannot read well. As they grow older, the size of their feet and their reading abilities increase, so there is a strong correlation between foot size and reading ability. Therefore, big feet cause people to read well.

b. Students who do homework learn what they are supposed to learn for a class. There is a strong correlation between the amount of homework students do and the grades they get in a class. Therefore, doing homework causes students to get good grades.

c. *"Get a good education and you'll get a good job. . . . a CNI study of U.S. Census data suggests that there is a direct correlation between education and income levels. The suburbs with the wealthiest population and the highest property values are also the ones with the highest percentage of college graduates . . . . the community that has the most adults without a high school diploma is West Milwaukee . . . ranks dead last in median household income and median home value."*

Source: *Tri City Hub,* October 21, 1993.

Does having an education cause you to have a higher income?

d. In a certain game, there was a high positive correlation between the number of field goals basketball players made and the number of fouls they made. Therefore, if you would like to make a lot of points, you should make a lot of fouls.

SUMMARY

Correlation is a measure of the strength of the linear association between two variables.

- Correlation is always between 1 and $^-1$: $^-1 \le r \le 1$.
- If the pattern in a scatter plot is close to a straight line, the correlation will be ***strong***, and the absolute value of r will be close to 1.
- If the pattern in a scatter plot is scattered and not a straight line, or if the pattern is not linear, the correlation is ***weak*** and the value of r is close to 0.
- The scales used in a scatter plot do not affect the correlation.
- Correlation is only a measure of linear association. It will not tell you about any other relationship that may exist between the variables.
- For a perfect positive correlation of 1 or a perfect negative correlation of $^-1$, if you know one variable, you can exactly predict the value of the other.
- A strong correlation between two variables does not necessarily indicate that there is a cause-and-effect relationship between the variables. There are often other variables that are involved in the association.

Practice and Applications

9. Sketch a scatter plot that you think reflects each of the following statements. Indicate either a positive, negative, or zero correlation for each.

a. The more hours you work, the more money you make.

b. The more money you spend, the less you save.

c. Some people work long hours and save a lot of money; some people work long hours and do not save a lot of money.

d. The faster you drive, the less time it takes to arrive at your destination.

e. The older you get, the less sleep you need.

f. The more hours you spend exercising, the more calories you burn.

g. Some people study for many hours and get good grades; some people study very little and get good grades. Some people study many hours and get poor grades.

10. The following data show the amount spent per pupil and the reading scores of seventh graders in each middle school.

Milwaukee Middle School Per Pupil Spending and Reading Scores

Middle School	Spending (per pupil)	Reading Scores
Fulton	$3,408	16
Robinson	3,306	36
Wright	3,274	46
Kosciuszko	3,049	33
Parkman	3,006	13
Edison	2,975	27
Roosevelt	2,970	54
Walker	2,787	35
Muir	2,783	30
Bell	2,710	34
Sholes	2,649	29
Burroughs	2,642	32
Webster	2,637	46
Audubon	2,609	52
Steuben	2,548	55
Fritsche	2,525	41
Morse	2,374	81

Source: data from *Milwaukee Journal*, May 25, 1992.

a. Plot the reading scores as a function of per-pupil spending (*spending, reading scores*), and find the correlation.

b. According to the *Milwaukee Journal*, ". . . the correlation isn't always there." What do you think this statement means?

11. Is there any association between the various statistics collected about baseball players and teams? If a team has a high number of hits, will it have a high number of runs batted in? Following are data for the 1993 American League Championship Season. Divide the following pairs of variables among your group members: (*at bats, runs batted in*), (*hits, runs batted in*), (*hits, home runs*), (*hits, batting averages*), (*at bats, home runs*), and (*home runs, runs batted in*).

American League Championship Season: 1993 Team Ratings

Team	Avg	AB	R	H	HR	RBI
New York	0.279	5,615	821	1,568	178	793
Toronto	0.279	5,579	847	1,556	159	796
Cleveland	0.275	5,619	790	1,547	141	747
Detroit	0.275	5,620	899	1,546	178	853
Texas	0.267	5,510	835	1,472	181	780
Baltimore	0.267	5,508	788	1,470	157	744
Chicago	0.265	5,483	776	1,454	182	731
Minnesota	0.264	5,601	693	1,480	121	642
Boston	0.264	5,496	686	1,451	114	644
Kansas City	0.263	5,522	675	1,455	125	641
Seattle	0.260	5,494	734	1,429	161	681
California	0.260	5,391	684	1,399	114	644
Milwaukee	0.258	5,525	733	1,426	125	688
Oakland	0.254	5,543	715	1,408	158	679

Source: data from *World Almanac and Book of Facts,* 1994.
(H = hit: batter gets on base; R = run: player scores; RBI = runs batted in: other runners scored on player's hit; HR = home run: batter scores on own hit; AB = at bat: batter has opportunity to bat; Avg = average: hits divided by at bats)

a. Make a scatter plot, find the correlation, and describe what it tells you about the association between the variables.

b. If you look at all the scatter plots, between which pair of variables is the association strongest? The weakest?

12. Federal Proposition 48 raised the academic requirements for participation in sports for athletes entering college in 1986. The college graduation rates in the Big Ten and the Pac Ten colleges for students who entered college in 1985 and 1986 are shown in the table shown on the next page. For example, in 1986, 78% of the students who entered the University of Illinois graduated, whereas only 74% of the athletes who entered that year graduated.

College Graduation Rates

College	Graduation Percentage of Students entering in 1986	Graduation Percentage of Athletes entering in 1985	Graduation Percentage of Athletes entering in 1986	Graduation Percentage of Football Players entering in 1985	Graduation Percentage of Football Players entering in 1986
Illinois	78	64	74	71	75
Indiana	65	66	62	72	50
Iowa	59	64	63	58	63
Michigan	85	63	79	61	80
Michigan State	69	65	62	48	56
Minnesota	42	53	53	42	32
Northwestern	89	82	77	75	62
Ohio State	54	67	69	63	72
Penn State	77	75	78	81	78
Purdue	70	61	65	47	55
Wisconsin	70	64	69	63	81
Arizona	49	50	54	53	57
Arizona State	45	40	52	24	30
California	77	62	61	57	44
Oregon	54	57	66	60	63
Oregon State	52	49	47	52	70
Southern Cal	66	53	69	60	71
Stanford	92	89	86	84	79
UCLA	74	58	60	55	42
Washington	63	56	61	35	54
Washington State	55	53	49	38	50

Source: data from *USA TODAY*, July 8, 1993.

a. Which college seems to have the worst record for athletes? The best?

b. Consider the four associations that follow. Divide the problems among your group members. Estimate the correlation using the tick-tack-toe method.

- Compare the percentage of all students who entered in 1986 and graduated from the college to the percentage of athletes who entered in 1986 and also graduated.
- Compare the percentage of all students who entered in 1986 and also graduated from the college to the percentage of football players who entered in 1986 and also graduated.

- Compare the percentage of football players who entered in 1985 and also graduated to the percentage of football players who entered in 1986 and also graduated.
- Compare the percentage of athletes who entered in 1985 and graduated to the percentage of all students who entered in 1986 and graduated.

c. What conclusions can you draw about the associations between the percentages of graduating athletes? Use your work from **b** to support your claims.

13. An article in the *Milwaukee Journal* in January 1993 reported the following:

Battle of bulge: Waist-hip ratio is predictor of death, study says

In a study of 41,837 Iowa women ages 55 to 69, researchers at the University of Minnesota found that there was a strong correlation between the waist-hip ratio and the risk of death. The bigger the waist in comparison to the hips, the higher the risk of death regardless of weight. A 15% increase in waist-hip ratio increased the risk of death by 60%.

The article continued by stating that for women, the ratio should be less than 0.8 and for men, it should be less than 0.95.

a. According to the article, who would be expected to have a greater risk of death: woman A, who has a waist measurement of 26 inches and a hip measurement of 36 inches; or woman B, who has a waist measurement of 32 inches and a hip measurement of 39 inches?

b. Make a sketch of a scatter plot that might show the research data.

c. How do you think the 15% increase in waist-hip ratio, that increased the risk of death by 60%, would be reflected in the scatter plot?

d. Based on the information in the article, which of the following do you think would be the most likely correlation for the data: $^-0.9$, $^-0.6$, $^-0.2$, 0, 0.2, 0.6, or 0.9? Tell why you selected the value you did.

e. Comment on the cause-and-effect relationship implied in the newspaper article.

End-of-Module Project: The Domino Topple

On January 2, 1988, thirty university students from Europe set up a continuous string of 1,500,000 dominoes. How long do you think it would take to set up that many dominoes?

Of that number, the students were able to topple (knock over) 1,382,101 dominoes with one push, setting a new world record according to the 1994 *Guinness Book of Records.* How long would it take you to topple a continuous string of 1,382,101 dominoes?

OBJECTIVE

Explore the relationship between several variables.

Suppose you would like to be able to predict with some certainty how long it will take a string of dominoes to fall over. Think about the rate of change and whether it will remain constant as the number of dominoes increases. Will the dominoes start falling faster or slower as more and more dominoes fall? Use what you have learned in this unit to help you find a solution.

INVESTIGATE

The Domino Topple

Devise a domino experiment of your own. Work with at least one other student. Be as complete as possible in your work.

Data Collection and Analysis

1. Define the variables in your problem. The only rule is that when a domino falls, it can knock over only one other domino.

2. Decide how you can best control these variables. Design an experiment using a box of dominoes, a meterstick, and a stopwatch.

3. Describe how you designed your experiment, what data you collected, and how you used the data to decide on your solution. You should have scatter plots and data tables showing how you reached your solution.

4. How will your solution change if

   a. You change the distance between the dominoes?

   b. You change any of your other variables?

LESSON 12 (OPTIONAL)

Catapults and Candy

How far will a gummy candy travel when it is launched from a certain height?

Is there a relationship between the height of the launch end of the pad and the distance the candy travels?

OBJECTIVES

Collect and organize data. Investigate the relationship between two variables. Use slope and intercept in meaningful contexts. Write an equation that can be used to make predictions. Design an experiment, and understand how the variables can change the results. Prepare an argument based on data.

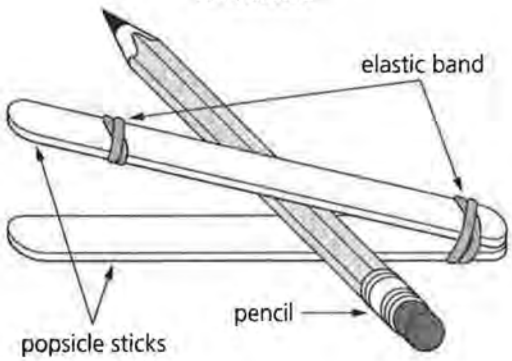
Modern catapults employ tension, hydraulic pressure, and other forces to launch airplanes from the decks of aircraft carriers. Follow the directions to make a catapult that uses tension, and experiment to answer the questions above.

INVESTIGATE

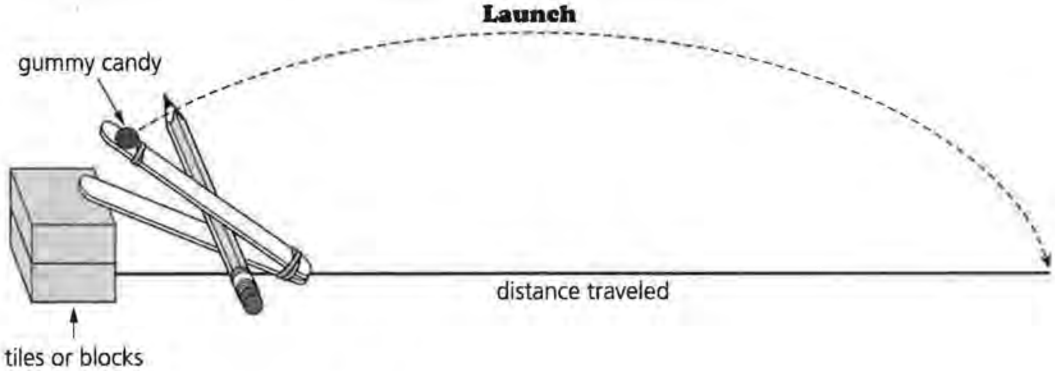
The Gummy Candy Launch

Each group will need at least one gummy candy (for example, a bear), about 10 tiles or small blocks, 2 ice-cream sticks, 2 elastic bands, a pencil, a meterstick, and a graphing calculator or graph paper. Each group should build a launchpad as shown here, place a gummy candy on the end of the launcher (an elastic band keeps the gummy candy from sliding down the ice-cream stick), and press down on the end of the launcher to release the candy. See the following diagrams of the launcher and the launch.

Launcher



Launch



Data Collection and Analysis

1. Work with at least one partner. Define the variables in your experiment and decide how you can control them.
2. Conduct an experiment to answer these questions:
   a. On average, how far will a gummy candy travel from a given height?
   b. How much variability is there in a given launch?
3. What is the relationship between the height of the launchpad and the distance the candy travels? Write an equation to describe the relationship. Explain your equation in terms of the data.
4. Use your answers to Question 3.
   a. If the height is 7 tiles, predict how far the candy will travel.
   b. If the height is 10 tiles, what is the expected distance the candy will travel?
   c. At what height should the launchpad be for the candy to travel 40 centimeters?

5. Examine your predictions.

   a. For how many values did your equation predict a greater distance than the experimental average distance? For which height was the prediction from the equation the worst?

   b. How confident are you in your equation? Explain.

Notes

Dale Seymour Publications® is a leading publisher of K–12 educational materials in mathematics, thinking skills, science, language arts, and art education.

90000

9 781572 322103

ISBN 1-57232-210-1

21161

DALE SEYMOUR PUBLICATIONS®