19.4 teacher notes

Objectives

- Make a scatter plot.
- Describe the correlation between two data sets.
- Write the equation for a line of fit.
- Use the equation of the line of fit to make predictions.

Vocabulary

Correlation
Line of fit
Negative correlation
Ordered pair (Lesson 7-1)
Point-slope form (Lesson 8-3)
Positive correlation
Quadrant (Lesson 7-1)
Regression line
Scatter plot
Slope (Lesson 7-2)
Slope-intercept form (Lesson 7-2)
Trend

Prerequisites

Graphing ordered pairs

Reading graphs

Writing the equation of a line given two points on the line

Evaluating expressions

Get Started

- Give each student a piece of large-grid graph paper and one single strand of spaghetti. Ask the students to draw *x* and *y* axes so that the graph paper shows the first quadrant of the coordinate plane. Students should label the *x*-axis "Number of Points Last Week" and the *y*-axis "Number of Points This Week."
- Write the ordered pairs (2, 2), (3, 3), (4, 2), (4, 5), (6, 3), (6, 5), (7, 7), (8, 5), and (9, 8) on the board. Students should then graph them. (You may want to prepare the graph with the points already plotted, to copy it, and to distribute it.) Each point on the graph shows the number of points scored last week and this week by an individual player. The point (5, 4) represents a player who scored five points last week and four points this week.
- Ask the class to describe the data. Ask, "What conclusions can you make based on the data?" Guide them to see that, in general, the more points a player scored last week, the more points she scored this week.

- Say, "The coach of the team would like to write a linear equation she can use to predict the number of points a player will score this week which are based on the number of points that player scored last week. Since the points don't lie exactly in a straight line but form a linear pattern, we'll write the equation of a line that we think represents the data fairly well. Use your piece of spaghetti to try to place such a line. Lay it on the graph so that no data point is too far from the line."
- As students experiment with the spaghetti, discuss their placement strategies. Students may offer varied suggestions on how to make the line, such as having an equal number of data points above and below the line or connecting the extreme data points.
- Say, "In today's lesson, we will learn about *lines of fit*. A line of fit is a line that fits the data fairly well. Lines of fit are important because they allow us to write functions which can be used to predict output values for given input values."

Section 1

Expand Their Horizons

In this lesson, students will learn how to draw and interpret a scatter plot, will find an equation of a line of fit, and will use the equation to make predictions. Section 1 will address drawing and interpreting a scatter plot.

Before beginning the lesson, review scatter plots. A scatter plot is a graph in which two data sets are compared by forming and graphing ordered pairs. Scatter plots are useful because they show whether there is a relationship between data sets.

In this lesson, only linear relationships are discussed. If the points of a scatter plot are near a line with positive slope (rising from left to right), the data sets have a *positive correlation*. As x increases, so does y. If the points of a scatter plot are near a line with negative slope (falling from left to right), the data sets have a *negative correlation*. If the points are scattered randomly, the data sets have *no correlation*.

It is important to note that a scatter plot does not necessarily represent a function. That is, the same input value may be associated with different output values. For example, in one movie that lasts $1\frac{1}{2}$ hours, Ferd eats two bags of gummi-slugs; in another movie that lasts $1\frac{1}{2}$ hours, he eats four bags. By analogy to a functional relationship in x and y, the independent variable is on the horizontal (x) axis and the dependent variable is on the vertical (y) axis in the scatter plot. So, in the problem about movie revenue (sales), the number of weeks since the movie's release is the independent variable, and the sales is the dependent variable.

Use the data below to answer Guided Notes Questions 1 and 2.

Temp. (°F)	People at the beach
62	35
95	90
90	60
70	55
85	80
87	70

1 **Make a scatter plot of the data.** Point out that it is reasonable to presume that the number of people at the beach depends on the temperature, so the number of people

at the beach is represented by the dependent variable, and temperature is represented by the independent variable.

2 **Describe the relationship in the scatter plot which compares the number of people at a beach and the daily high temperature.** Because the number of people at the beach increases as the temperature increases, the points generally rise from left to right, and there is a positive correlation between the two data sets.

Section 2

Expand Their Horizons

In this section, lines of fit will be given for several data sets. Then, the task is to find equations for the lines and use the equations to predict values that are not in the original data sets. The name of the lesson is "Finding a Line of Best Fit," but the lesson actually does not address how to find **the line of best fit**, because the method required for that is beyond the scope of this course. Instead, the lesson uses a line of fit, which is not necessarily the line of best fit. See the "Look Beyond" section for a description of a method for finding the line of best fit.

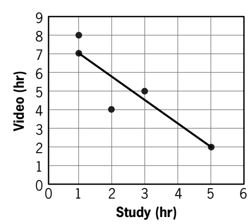
If there is a positive correlation between data sets, a line of fit for the scatter plot has a positive slope, and if there is a positive correlation between data sets, a line of fit for the scatter plot has a positive slope. A line of fit may or may not actually contain any of the data points. For most lines of fit, there should be about the same number data of points above the line as below. An exception to this rule of thumb is a case in which one or more data points are significantly farther from the line than most of the other points. In such a case, there may be fewer data points on the side of the line with the data points that are farther away.

If two points on a line can be identified or one point and the slope, then an equation of the line can be written. The equation can then be used to find, or to predict, the value of one variable when given a value of the other variable.

Review the slope formula. For a line containing points (x_1, y_1) and (x_2, y_2), slope $m = \frac{y_2 - y_1}{x_2 - x_1}$. Slope is a numerical measure of the "tilt" of a line. A line with positive slope rises from left to right; a line with negative slope falls from left to right. A line with zero slope is horizontal.

To write an equation of a line when two points on the line are known, begin by finding the line's slope. Then, use the point-slope form $y - y_1 = m(x - x_1)$ to write the equation of the line, using either of the known points as (x_1, y_1). Finally, use the properties of equality to write the equation in the form $y = mx + b$ (slope-intercept form).

Use the scatter plot below to answer Guided Notes Questions 3 and 4.



3 **Find the equation for the line of fit shown on the scatter plot which compares the number of hours studying and the number of hours playing video games per week for five weeks. Use the points (1, 7) and (5, 2) to find the equation.** Using the slope formula, $m = \frac{2 - 7}{5 - 1} = \frac{-5}{4} = -1.25$. Using (1, 7) as (x_1, y_1) in the point-slope form, $y - 7 = -1.25(x - 1)$. On the right side of the equation, distribute –1.25 to get $y - 7 = -1.25x + 1.25$. Write the equation in slope-intercept form by adding seven to both sides. The equation of the given line of fit is $y = -1.25x + 8.25$. Note the same equation will be found if (5, 2) is used for (x_1, y_1). Begin with $y - 2 = -1.25(x - 5)$, distribute –1.25 to get $y - 2 =$ $-1.25x + 6.25$, and then, add two to both sides to get $y = -1.25x + 8.25$.

Look Beyond

The purpose of a line of fit is to represent a pattern of data points. In this lesson, a line of fit is placed so that roughly half the data points lie below the line and half above. This method of placing a line of fit is easy to understand, but it is not likely to result in the line of best fit. To find the line of best fit, a statistical method called the *least squares method* is used. In the least squares method, a line is found so that the sum of the squares of the vertical distances from the data points to the line is minimized. The least squares method can be accomplished with paper and pencil, but for practical purposes, a scientific calculator or computer spread sheet program is most often used.

The equation of a line of best fit is called a linear regression equation. Sometimes, the data points in a scatter plot are in a nonlinear pattern. Best-fit equations can be found for nonlinear patterns also; two of the more common types are quadratic and exponential regression equations.

4 **Predict the number of hours playing video games if the student spends three hours studying.** Recall in Guided Notes Question 3, x represented the number of hours spent studying and y represented the number of hours spent playing video games. So, if the student spends three hours studying, he will spend $-1.25(3) + 8.25$ or 4.5 hours playing video games.

Connections

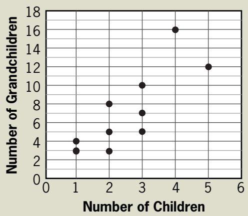
Linear regression analysis (finding the equation of the line of best fit) is often used in scientific research. For example, United States Geological Survey researchers used linear regression analysis to study the relationship between the amount of water in wetlands and the number of nests of two types of waterfowl. The researchers conducted their study in a particular location in North Dakota. They made a scatter plot of the percentage of wetlands containing water and the number of nests built in those wetlands by mallards and northern pintails from 1966 to 1981. The scatter plot indicated a positive correlation, and scientists were able to write a linear equation to express the relationship. The equation can be used to predict the number of nests which will be built during given water conditions.

Additional Examples

1. Ten grandmothers were interviewed and asked for the number of children and grandchildren they have. The results are shown in the table.

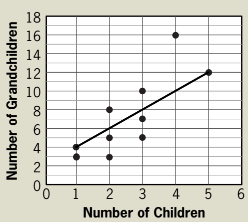
Number of Children	Number of Grandchildren	Number of Children	Number of Grandchildren
4	16	2	3
3	7	2	5
5	12	1	4
2	8	3	10
1	3	3	5

Using number of children as the independent variable and number of grandchildren as the dependent variable, make a scatter plot of the data and describe the relationship.



There is a **positive** correlation between the number of children and the number of grandchildren.

2. Find an equation for the line of fit shown below. Then, use the equation to predict the number of grandchildren a grandmother with six children may have.



The line contains the points (1, 4) and (5, 12). Students may notice that five points lie below the line, and three points lie above it. This is justified because the point (4, 16) is significantly farther from the line than any of the other points.

Slope Formula:

$$m = \frac{y_2 - y_1}{x_2 - x_1} = \frac{12 - 4}{5 - 1} = \frac{8}{4} = 2$$

Point-Slope Form:

$$y - y_1 = m(x - x_1)$$
$$y - 4 = 2(x - 1)$$
$$y - 4 = 2x - 2$$
$$y - 4 + 4 = 2x - 2 + 4$$
$$y = 2x + 2$$

Prediction:

$$y = 2x + 2$$
$$y = 2(6) + 2$$
$$y = 14$$

Using the equation of the line of fit, we can predict that a grandmother with six children will have 14 grandchildren.

Manipulatives

A geoboard can be used to help students find a line of fit. A 10×10 or 11×11 geoboard is preferable to a 6×6 geoboard for this activity. If only a 6×6 geoboard is available, adjust the data points in the activity to accommodate the smaller board.

To plot points on a geoboard, use paper reinforcements (rings of gummed paper used for reinforcing the holes punched in paper). Using the data from Guided Practice Question 1, students should plot the data points on the geoboard. Remind them that the bottom-most row and left-most column of the geoboard represent the x-axis and y-axis, respectively. Students should use rubber bands to form the axes. When all the data points have been plotted, ask the class to experiment with a rubber band to find a good line of fit. After making a "line" by placing the rubber band around two pegs, students should then study how well the line represents the data points. Move around the room and ask the following questions: "Are there generally about as many points above the line of fit as below it? Is each data point about the same distance from the line of fit? Is any one data point extremely far from the line of fit? If so, how can the line of fit be adjusted?" Encourage the class to agree on the best line of fit, identifying it by two points it contains.

Calculator Problem

Finding the line of best fit requires complex statistical computations beyond the scope of an Algebra I course. However, many graphing calculators can perform the task. The instructions given in this sample refer to the TI-83 graphing calculator. For instructions on finding the line of best fit on another graphing calculator, see the owner's manual.

To determine the equation of the line of best fit in the lesson example about Ferd's consumption of gummi-slugs, begin by entering the data into the calculator through the STAT **1:Edit...** menu. See Figure 1. Next, make a scatter plot of the data. Select 2nd Y= . Select **1:Plot1...Off**. Select **On** and **Type**: scatter plot. Now select GRAPH. See Figure 2. To find the line of best fit, use the calculator's linear regression function. Select STAT, **CALC**, **4:LinReg(ax+b)**, ENTER, ENTER. The graphing calculator produces a line of best fit with equation $y = 2.303030303x + 0.189393939$. See Figure 3.

Figure 1

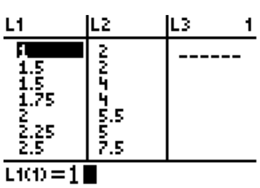


Figure 2

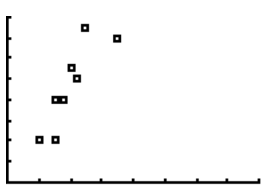
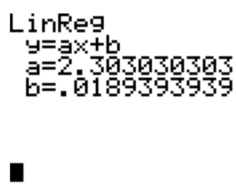


Figure 3



Computer/Spreadsheet

An automated spreadsheet program, like Microsoft® Excel, can also be used to find the line of best fit. Begin by entering ordered pairs so that x-values appear in Column A and y-values in Column B. See Figure 4.

Then, insert a chart of the data. Select the numeric data. Go to Insert, Chart then select XY(Scatter). See Figure 5. Now, add a trendline to show the linear regression model. Click on the plot area. Now select Chart, Add Trendline, OK. See Figure 6. The graph now shows the scatter plot and the line of best fit. The TI-83 and Microsoft® Excel will produce the same line of best fit.

Figure 4

A	B
1	2
1.5	2
1.5	4
1.75	4
2	5.5
2.25	5
2.5	7.5
3.5	7

Figure 5

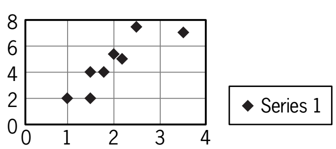
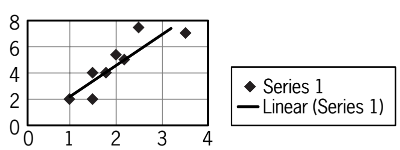


Figure 6



After completing the graphing calculator and/or spreadsheet activity, discuss with the class how the calculator's results compare to their own (in the lesson the line of fit they found was $y = 2x + 0.5$). The slopes, 2 and 2.30, are relatively close. A slope of two indicates that for each hour of movie length, Ferd can expect to eat about two boxes of gummi-slugs. The y-intercepts of 0.5 and 0.0189 are more disparate. Logically, a y-intercept of 0.0189 (which is nearly 0) makes more sense, because it is reasonable to assume that Ferd will eat no gummi-slugs in a zero-hour movie.