Name _____

Lesson 19: Interpreting Correlation

Warm Up: Regents Prep

LEARNING OUTCOMES (L.O.)

- I can compute and interpret the correlation coefficient of a linear fit.
- I can describe the difference between correlation and causation.

1. In the equation $x^2 + 10x + 24 = (x + a)(x + b)$, b is an integer.

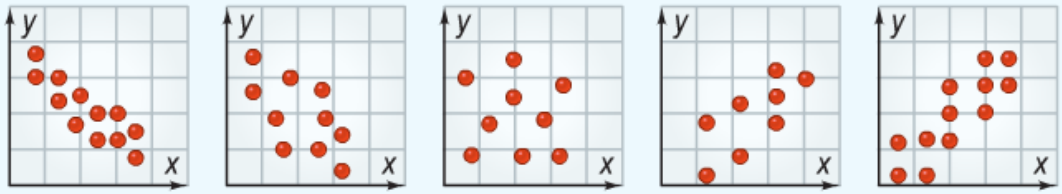
   Find algebraically *all* possible values of b.

2. Rhonda deposited $3000 in an account in the Merrick National Bank, earning 4.2% interest, compounded annually. She made no deposits or withdrawals. Write an equation that can be used to find B, her account balance after t years.

A **scatter plot** is a graph that relates two sets of data by plotting the data as ordered pairs. You can use a scatter plot to determine the strength of the relationship, or **correlation**, between data sets. The closer the data points fall along a line with positive slope,

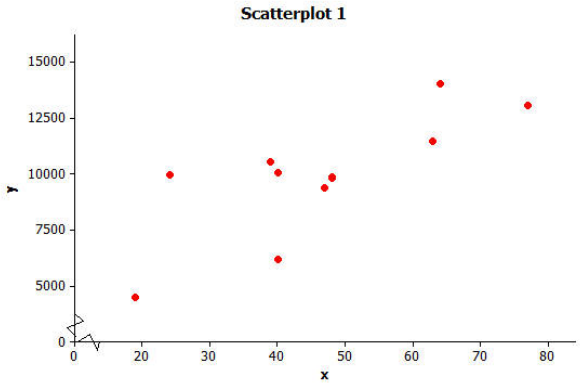
- the stronger the linear relationship and
- the stronger the positive correlation

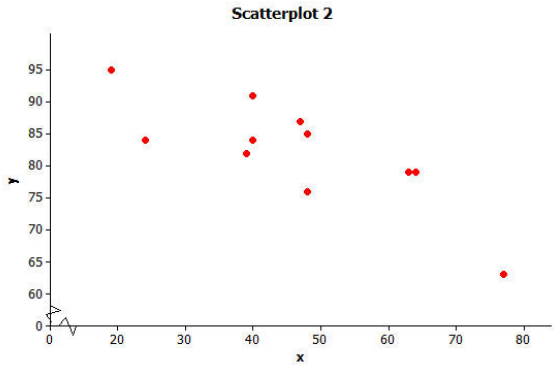
between the two variables.



Positive and Negative Linear Relationships

Linear relationships can be described as either positive or negative. Below are two scatter plots that display a linear relationship between two numerical variables x and y.





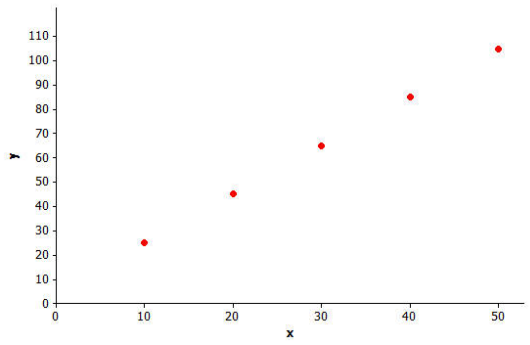
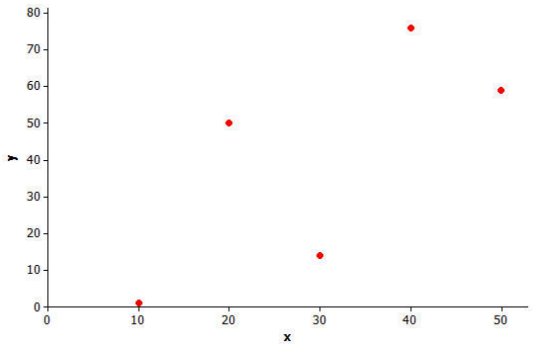
1. The relationship displayed in Scatter Plot 1 is a positive linear relationship. Does the value of the y variable tend to increase or decrease as the value of x increases? If you were to describe this relationship using a line, would the line have a positive or negative slope?

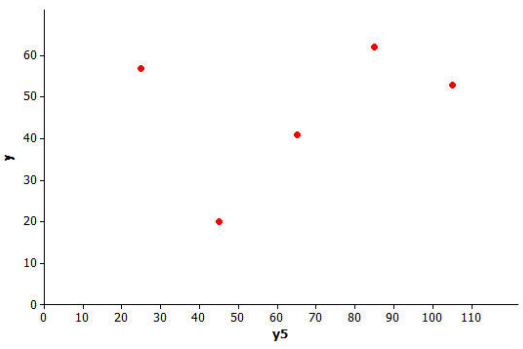
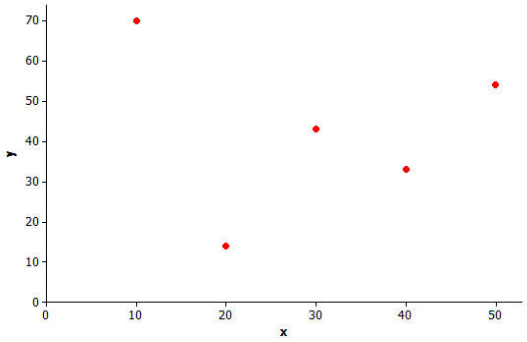
2. The relationship displayed in Scatter Plot 2 is a negative linear relationship. As the value of one of the variables increases, what happens to the value of the other variable? If you were to describe this relationship using a line, would the line have a positive or negative slope?

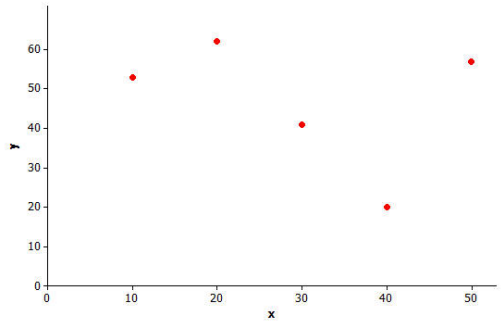
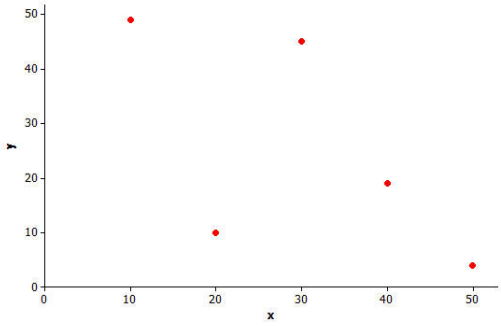
The Correlation Coefficient

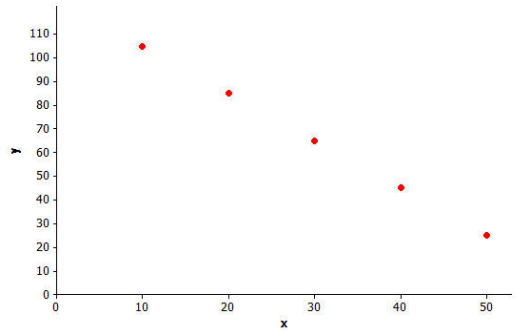
The correlation coefficient is a number between −1 and +1 (including −1 and +1) that measures the strength and direction of a linear relationship. The correlation coefficient is denoted by the letter r.

Several scatter plots are shown below. The value of the correlation coefficient for the data displayed in each plot is also given.

	*r = 1.00*
	*r = 0.71*

	*r = 0.32*
	*r = −0.10*

	*r = −0.32*
	*r = −0.63*

	*r = −1.00*

3. When is the value of the correlation coefficient positive?

4. When is the value of the correlation coefficient negative?

5. Is the linear relationship stronger when the correlation coefficient is closer to 0 or to 1 (or –1)?

Here are some important properties of the correlation coefficient:

Property 1: The sign of r (positive or negative) corresponds to the direction of the linear relationship

Property 2: A value of $r = +1$ indicates a perfect positive linear relationship, with all points in the scatter plot falling exactly on a straight line.

Property 3: A value of $r = -1$ indicates a perfect negative linear relationship, with all points in the scatter plot falling exactly on a straight line.

Property 4: The closer the value of r is to $+1$ or -1, the stronger the linear relationship.

Calculating the Value of the Correlation Coefficient

There is an equation that can be used to calculate the value of the correlation coefficient given data on two numerical variables. Using this formula requires a lot of tedious calculations that will be discussed in later grades. Fortunately, a graphing calculator can be used to find the value of the correlation coefficient once you have entered the data.

Your teacher will show you how to enter data and how to use a graphing calculator to obtain the value of the correlation coefficient.

Here is the data from a previous lesson on shoe length in inches and height in inches for 10 men.

Shoe Length (*x*)	Height (*y*)
inches	inches
12.6	74
11.8	65
12.2	71
11.6	67
12.2	69
11.4	68
12.8	70
12.2	69
12.6	72
11.8	71

6. Enter the shoe length and height data in your calculator. Find the value of the correlation coefficient between shoe length and height. Round to the nearest tenth.

The table below shows how you can informally interpret the value of a correlation coefficient.

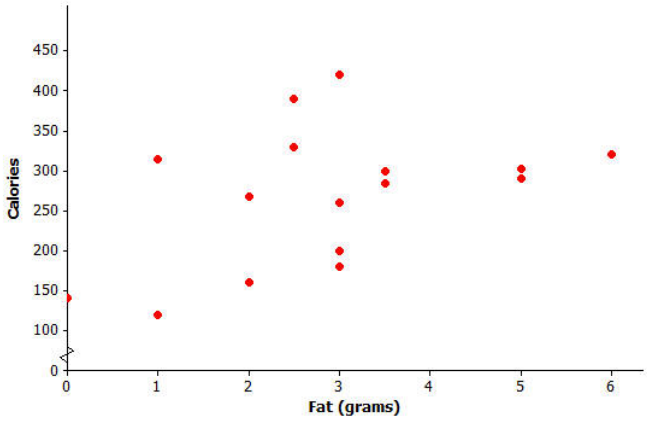
If the value of the correlation coefficient is between…	You can say that…
$r = 1.0$	There is a perfect positive linear relationship.
$0.7 \leq r < 1.0$	There is a strong positive linear relationship.
$0.3 \leq r < 0.7$	There is a moderate positive linear relationship.
$0 < r < 0.3$	There is a weak positive linear relationship.
$r = 0$	There is no linear relationship.
$-0.3 < r < 0$	There is a weak negative linear relationship.
$-0.7 < r \leq -0.3$	There is a moderate negative linear relationship.
$-1.0 < r \leq -0.7$	There is a strong negative linear relationship.
$r = -1.0$	There is a perfect negative linear relationship.

7. Interpret the value of the correlation coefficient between shoe length and height for the data in the height versus shoe length example.

Practice Calculating and Interpreting Correlation Coefficients

Consumer Reports published a study of fast-food items. The table and scatter plot below display the fat content (in grams) and number of calories per serving for 16 fast-food items.

Fat (g)	Calories (kcal)
2	268
5	303
3	260
3.5	300
1	315
2	160
3	200
6	320
3	420
5	290
3.5	285
2.5	390
0	140
2.5	330
1	120
3	180



Data Source: *Consumer Reports*

8. Based on the scatter plot, do you think that the value of the correlation coefficient between fat content and calories per serving will be positive or negative? Explain why you made this choice.

9. Based on the scatter plot, estimate the value of the correlation coefficient between fat content and calories.

10. Calculate the value of the correlation coefficient between fat content and calories per serving. Round to the nearest hundredth. Interpret this value.

Correlation Does Not Mean There is a Cause-and-Effect Relationship Between Variables

It is sometimes tempting to conclude that if there is a strong linear relationship between two variables that one variable is causing the value of the other variable to increase or decrease. But you should avoid making this mistake. When there is a strong linear relationship, it means that the two variables tend to vary together in a predictable way, which might be due to something other than a cause-and-effect relationship.

For example, the value of the correlation coefficient between sodium content and number of calories for the fast food items in the previous example was $r = 0.79$, indicating a strong positive relationship. This means that the items with higher sodium content tend to have a higher number of calories. But the high number of calories is not caused by the high sodium content. In fact sodium does not have any calories. What may be happening is that food items with high sodium content also may be the items that are high in sugar and/or fat, and this is the reason for the higher number of calories in these items.

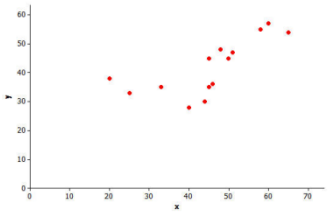
Similarly, there is a strong positive correlation between shoe size and reading ability in children. But it would be silly to think that having big feet causes children to read better. It just means that the two variables vary together in a predictable way. Can you think of a reason that might explain why children with larger feet also tend to score higher on reading tests?

Name _____ Date _____

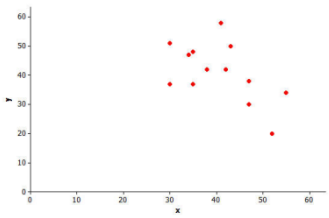
Lesson 19: Interpreting Correlation CW/HW

1. Which of the three scatter plots below shows the strongest linear relationship? Which shows the weakest linear relationship?

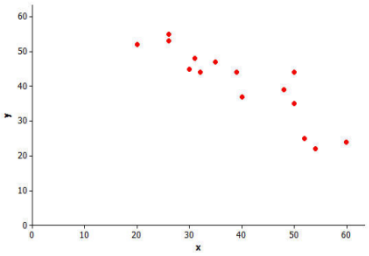
Scatter plot 1



Scatter plot 2



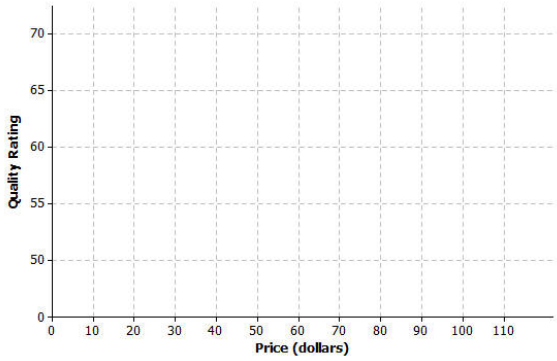
Scatter plot 3



2. *Consumer Reports* published data on the price (in dollars) and quality rating (on a scale of 0 to 100) for 10 different brands of men's athletic shoes.

Price ($)	Quality Rating
65	71
45	70
45	62
80	59
110	58
110	57
30	56
80	52
110	51
70	51

a. Construct a scatter plot of these data using the following grid.

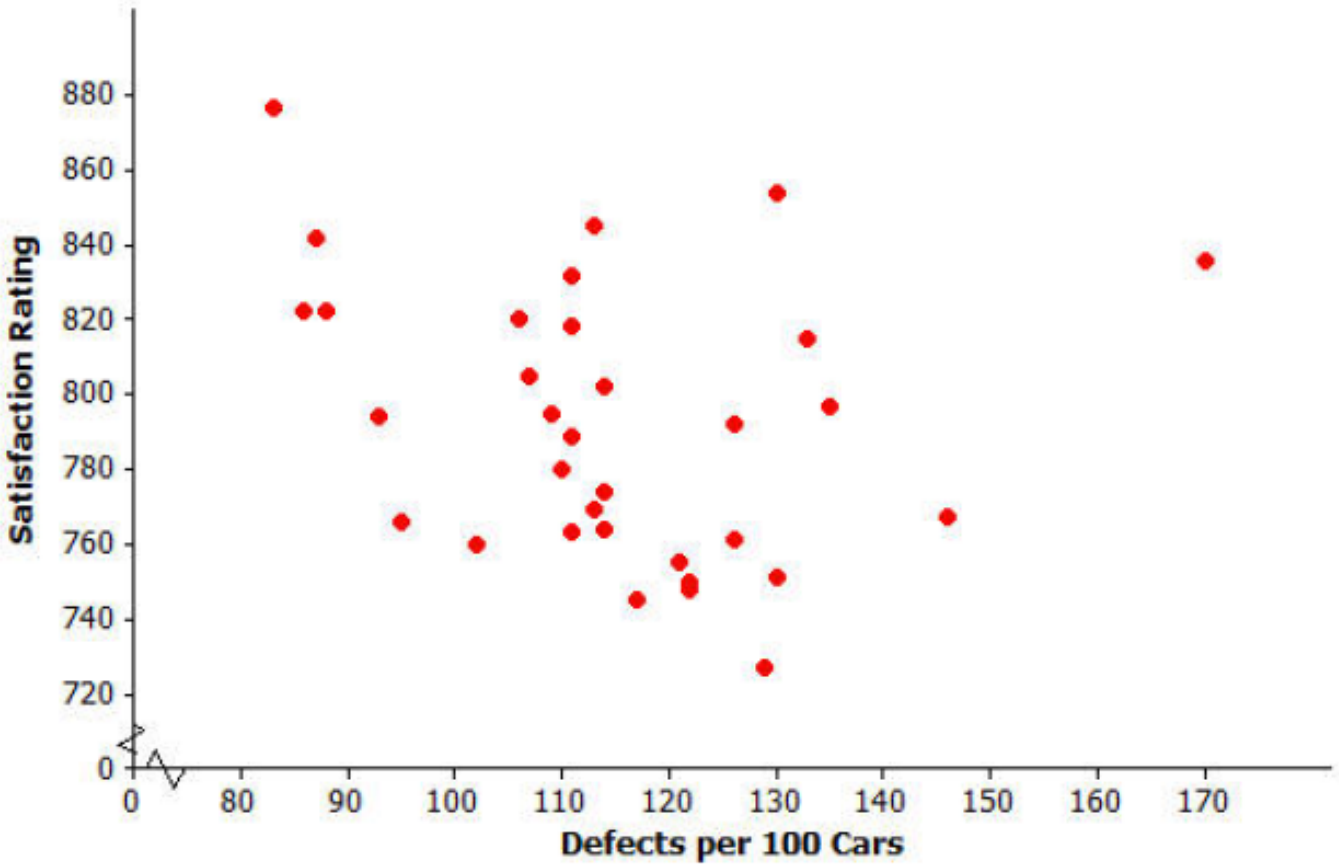


b. Calculate the value of the correlation coefficient between price and quality rating and interpret this value. Round to the nearest hundredth.

Name _____

Lesson 19: Interpreting Correlation
Exit Ticket

The scatter plot below displays data on the number of defects per 100 cars and a measure of customer satisfaction (on a scale from 1 to 1000, with higher scores indicating greater satisfaction) for the 33 brands of cars sold in the United States in 2009.



Data Source: *USA Today*, June 16, 2010 and July 17, 2010

Which of the following is the value of the correlation coefficient for this data set:

$r = -0.95$, $r = -0.24$, $r = 0.83$, or $r = 1.00$?

b. Explain why you selected this value.