Unit 3, Lesson 8

Using the Correlation Coefficient

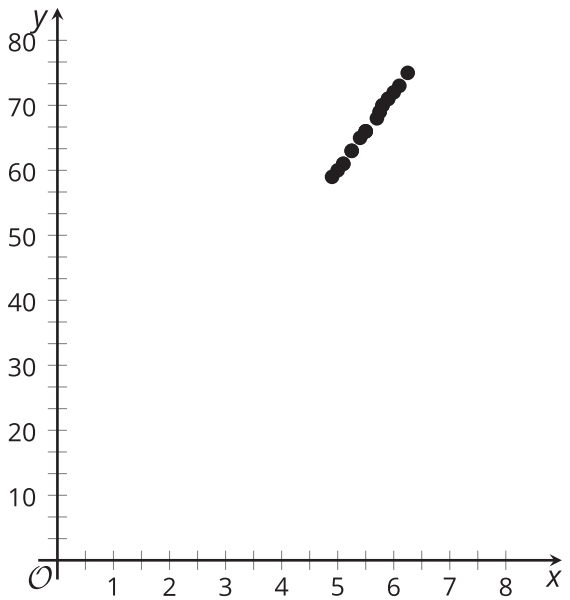
Let's look closer at correlation coefficients.

8.1 Putting the Numbers in Context

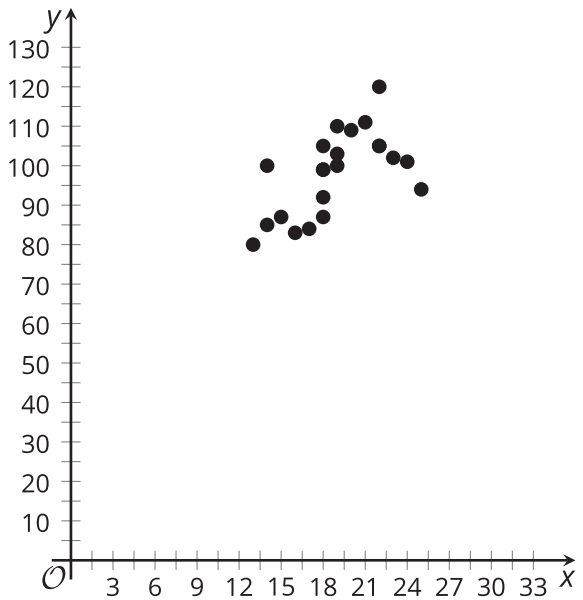
Match the variables to the scatter plot you think they best fit. Be prepared to explain your reasoning.

	x-variable	y-variable
1.	low temperature in Celsius for Denver, CO, on a given day	boxes of cereal in stock at a grocery store in Miami, FL, on a given day
2.	number of free throws shot in a game	basketball team score in a game
3.	measured student height in feet	measured student height in inches
4.	number of minutes spent in a waiting room	hospital satisfaction rating given by patient

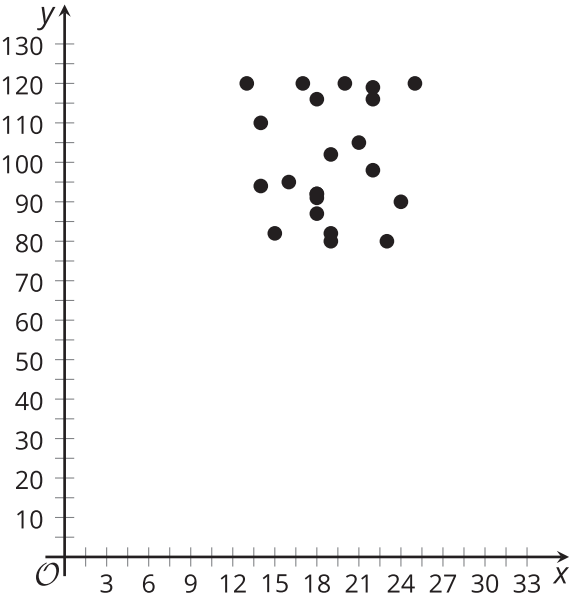
A


y
80
70
60
50
40
30
20
10
O
1 2 3 4 5 6 7 8
x

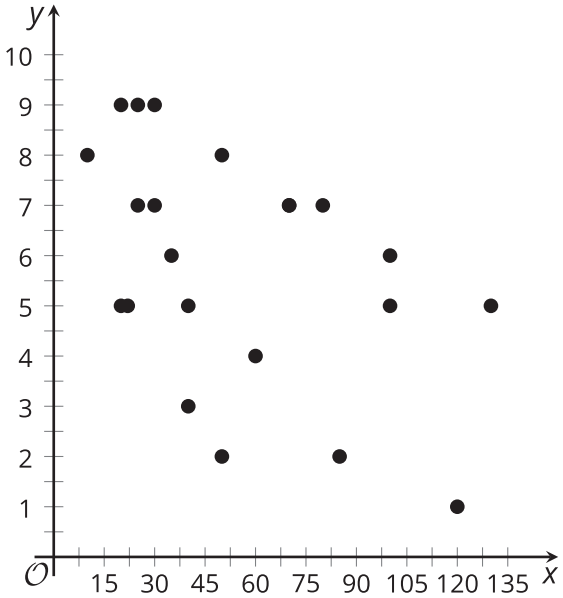
B


y
130
120
110
100
90
80
70
60
50
40
30
20
10
O
3 6 9 12 15 18 21 24 27 30 33
x

C


y
130
120
110
100
90
80
70
60
50
40
30
20
10
O
3 6 9 12 15 18 21 24 27 30 33
x

D


y
10
9
8
7
6
5
4
3
2
1
O
15 30 45 60 75 90 105 120 135
x

8.2 Never Know How Far You'll Go

Priya takes note of the distance the car is driven and the time it takes to get to the destination for many trips.

distance (mi) (x)	travel time (min) (y)
2	4
5	7
10	11
10	15
12	16
15	22
20	23
25	25
26	28
30	36
32	35
40	37
50	51
65	70
78	72

1. Distance is one factor that influences the travel time of Priya's car trips. What are some other factors?

2. Which of these factors (including distance) most likely has the most consistent influence for all the car trips? Explain your reasoning.

3. Use technology to create a scatter plot of the data and add the best fit line to the graph.
4. What do the slope and y-intercept for the line of best fit mean in this situation?

5. Use technology to find the correlation coefficient for this data. Based on the value, how would you describe the strength of the linear relationship?

6. How long do you think it will take Priya to make a trip of 90 miles if the linear relationship continues? If she drives 90 miles, do you think the prediction you made will be close to the actual value? Explain your reasoning.

8.3 Correlation Zoo

For each situation, describe the relationship between the variables, based on the correlation coefficient. Make sure to mention whether there is a **strong relationship** or **weak relationship** and whether it is a **positive relationship** or **negative relationship**.

1. Number of steps taken per day and number of kilometers walked per day. $r = 0.92$

2. Temperature of a rubber band and distance the rubber band can stretch. $r = 0.84$

3. Car weight and distance traveled using a full tank of gas. $r = -0.86$

4. Average fat intake per citizen of a country and average cancer rate of a country. $r = 0.73$

5. Score on science exam and number of words written on the essay question. $r = 0.28$

6. Average time spent listening to music per day and average time spent watching TV per day. $r = -0.17$

Are you ready for more?

A biologist is trying to determine if a group of dolphins is a new species of dolphin or if it is a new group of individuals within the same species of dolphin. The biologist measures the width (in millimeters) of the largest part of the skull, zygomatic width, and the length (in millimeters) of the snout, rostral length, of 10 dolphins from the same group of individuals.

The data appears to be linear. The equation of the line of best fit is $y = 0.201x + 110.806$, and the r-value is 0.201.

x, rostral length (mm)	y, zygomatic width (mm)
288	147
247	147
268	171
278	177
258	168
272	184
272	161
258	159
273	168
277	166

1. After checking the data, the biologist realizes that the first zygomatic width listed as 147 mm is an error. It is supposed to be 180 mm. Use technology to find the equation of a line of best fit and the correlation coefficient for the corrected data. What is the equation of the line of best fit, and what is the correlation coefficient?

2. Compare the new equation of the line of best fit with the original. What impact did changing 1 data point have on the slope, y-intercept, and correlation coefficient of the line of best fit?

3. Why do you think that weak, positive association became a moderately strong association? Explain your reasoning.

4. Use technology to change the y-value for the first and second entries in the table.

   a. How does changing each point's y-value impact the correlation coefficient?

   b. Can you change two values to get the correlation coefficient closer to 1? Use data to support your answer.

   c. By leaving $(288, 180)$, can you change a value to get the relationship to change from a positive one to a negative one? Use data to support or refute your answer.

Lesson 8 Summary

The value for the correlation coefficient can be used to determine the strength of the relationship between the two variables represented in the data.

In general, when the variables increase together, we can say they have a **positive relationship**. If an increase in one variable's data tends to be paired with a decrease in the other variable's data, the variables have a **negative relationship**. When the data is tightly clustered around the best fit line, we say there is a **strong relationship**. When the data is loosely spread around the best fit line, we say there is a **weak relationship**.

A correlation coefficient with a value near 1 suggests a strong, positive relationship between the variables. This means that most of the data tends to be tightly clustered around a line and that when one of the variables increases in value, the other does as well. The number of schools in a community and the population of the community is an example of variables that have a strong, positive correlation. When there is a large population, there is usually a large number of schools, and small communities tend to have fewer schools, so the correlation is positive. These variables are closely tied together, so the correlation is strong.

Similarly, a correlation coefficient near -1 suggests a strong, negative relationship between the variables. Again, most of the data tend to be tightly clustered around a line, but now, when one value increases, the other decreases. The variables time since leaving home and the distance left to reach school have a strong, negative correlation. As the travel time increases, the distance to school tends to decrease, so this is a negative correlation. The variables are again closely, linearly related, so this is a strong correlation.

Weaker correlations mean there may be other reasons the data is variable other than the connection between the two variables. For example, variables number of pets and number of siblings have a weak correlation. There may be some relationship, but there are many factors that account for the variability in the number of pets other than the number of siblings.

The context of the situation should be considered when determining whether the correlation value is strong or weak. In physics, where things are measured with precise instruments, a correlation coefficient of 0.8 may not be considered strong. In social sciences, where data is collected through surveys, a correlation coefficient of 0.8 may be very strong.