

Module 2: Lesson 19

Interpreting Correlation

Lesson Summary

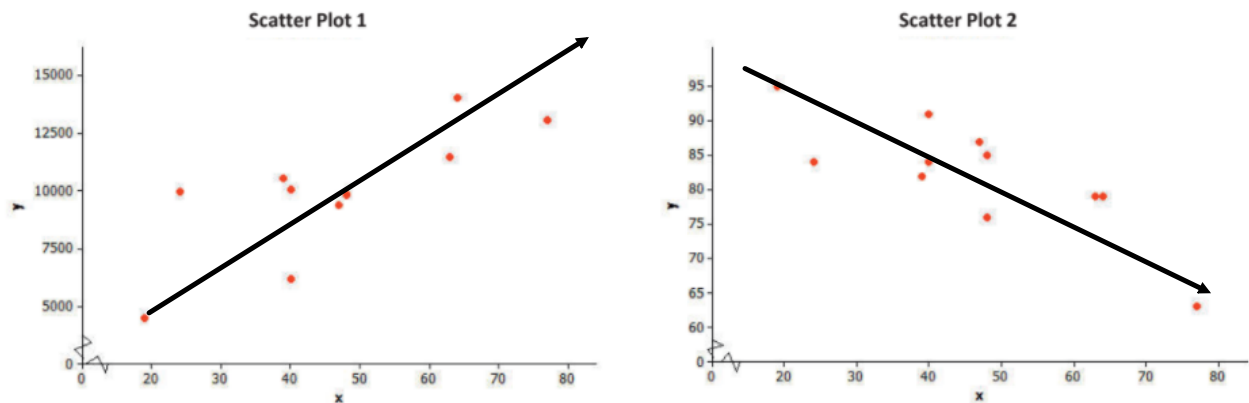
- Linear relationships are often described in terms of strength and direction.
- The correlation coefficient is a measure of the strength and direction of a linear relationship.
- The closer the value of the correlation coefficient is to $+1$ or -1 , the stronger the linear relationship.
- Just because there is a strong correlation between the two variables does not mean there is a cause-and-effect relationship.

Correlation is a statistical technique that is used to measure and describe the **STRENGTH** and **DIRECTION** of the relationship between two variables

Up until now we've just described correlation in a very general way such as "Negative Correlation", "Positive Correlation", and "No Correlation", all of which were based on how a relationship looks.

Starting in this section we will not only tell whether or not two variables are correlated, but we will determine which ones are **MORE CORRELATED** than others by assigning the correlation a value.

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?

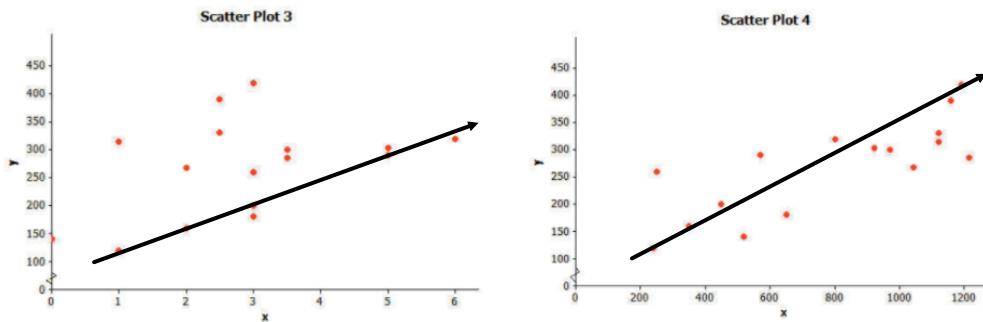
Scatter plot 1 has a positive slope.

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?

Scatter plot 2 has a negative slope.

Note: A linear correlation is "stronger" if the points are clustered together to more closely resemble a line.

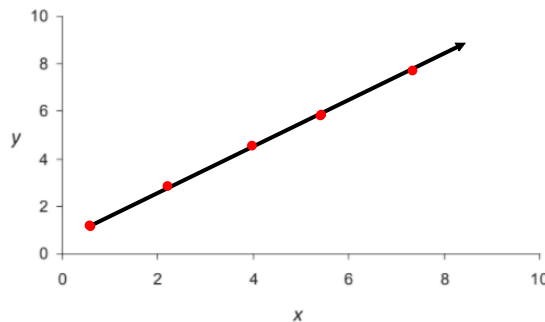
Below are two scatter plots that show a linear relationship between two numerical variables x and y .



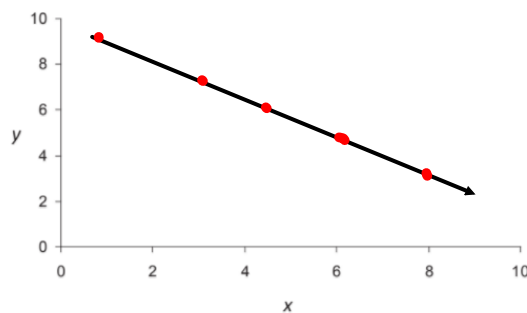
Which scatter plot has a stronger correlation than the other? Explain your reasoning.

Scatter Plot 4 has a stronger correlation

8. What do you think a scatter plot with the strongest possible linear relationship might look like if it is a positive relationship? Draw a scatter plot with five points that illustrates this.

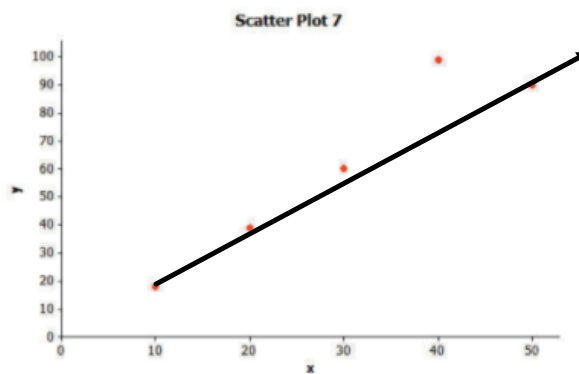
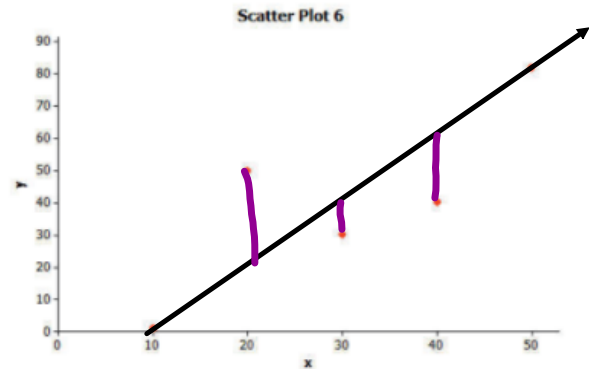
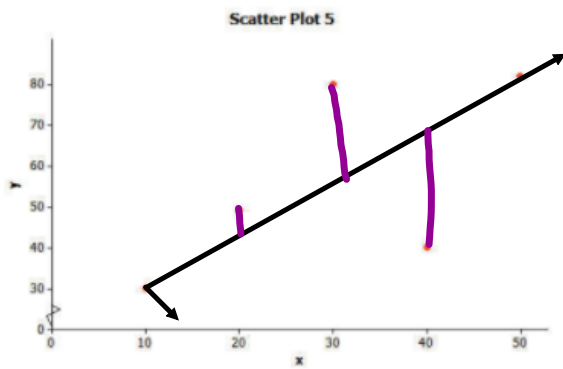


9. How would a scatter plot that shows the strongest possible linear relationship that is negative look different from the scatter plot that you drew in the previous question?



10. Consider the three scatter plots below. Place them in order from the one that shows the strongest linear relationship to the one that shows the weakest linear relationship.

Strongest	➔	Weakest
7	6	5



Correlation Coefficient

We've seen scatter plots with positive correlations and negative correlations. After that we determined, by just observing the patterns of each scatter plot, which graphs had stronger correlations than others. Now we're going to take this one step further by assigning a value to the correlation. This is called the "Correlation Coefficient" of a scatter plot.

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 for the data is denoted by the letter *r*.

$$r = \frac{\sum (x - \bar{x})}{\sqrt{\sum (x - \bar{x})^2} \sqrt{\sum (y - \bar{y})^2}}$$

Equation
for
correlation

But we will not use the above equation in this class

The following are examples of correlation coefficients with a given value and what their plot might look like.

	$r = 1.00$
	$r = 0.71$
	$r = 0.32$
	$r = -0.10$
	$r = -0.32$
	$r = -0.63$
	$r = -1.00$

From these scatter plots there are 4 properties that we should keep in mind regarding the correlation coefficients of scatter plots:

- 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.

To better generalize correlation coefficients and their strength you can follow along with the following chart to help with your interpretation:

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.

Steps to calculate "r" (the correlation coefficient) with a TI-83/84 calculator.

1. **Stats** → Edit
2. Enter your data for x in L1 and your data for y in L2
3. **2nd** → **Mode**
4. Access the calculator's Catalog by entering
2nd → Catalog
5. Scroll down until you get to "DiagnosticOn" and hit "Enter"
6. After step 5 go to "Stats" and hit the right arrow key to access the "Calc" menu and select "LinReg(ax+b)" and hit "Enter"
7. Hit "Enter" again. "r" will have been calculated and that will be the correlation coefficient.