

Correlation

by Sophia



WHAT'S COVERED

This tutorial will introduce correlation and correlation coefficient. Our discussion breaks down as follows:

1. Correlation
2. Correlation Coefficients and Scatter Plots

1. Correlation

When first describing scatter plots, you learned about their form, direction, and strength.

- *Form* assessed the linearity
- *Direction* says whether the data points tend to move in a positive or negative direction
- *Strength* shows how well they follow that form.

When the form is linear, we can use a number called **correlation coefficient**. It measures the strength and direction of a linear relationship. The direction will be easy to spot. It will be a positive number if there's a positive association, and a negative number if there's a negative association. The numerical quantity will measure strength. The correlation coefficient is a variable called " r " and is unit-less. It is expressed as a number between negative 1 and positive 1 and indicates the strength of the linear association.

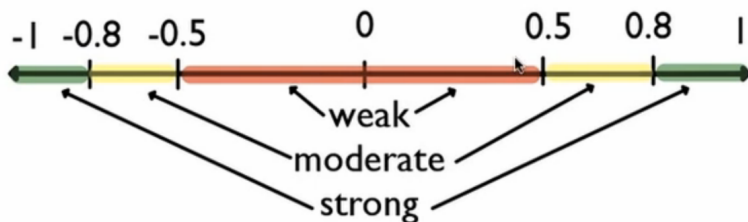


BIG IDEA

In statistical terms, correlation describes the relationship between two variables. When the explanatory variable and the response variable both increase together, we refer to this as a **positive correlation**.

Conversely, if one variable decreases as the other increases, it is known as a **negative correlation**.

Numbers that are close to negative 1 or positive 1 are associated with a strong association between the two variables—a 1 indicating a strong positive association, and a negative 1 indicating a strong negative association. Numbers near zero represent almost no linear relationship.



HINT

You can use the chart above to help you to understand the value of a correlation. Numbers between 0.8 and 1 are considered to have a strong correlation; between 0.5 and 0.8 a moderate correlation; and between 0 and 0.5 a very weak correlation. The same exists between negative 1 and 0.

TERMS TO KNOW

Correlation

The strength and direction of a linear association between two quantitative variables.

Correlation Coefficient

The numerical value between -1 and +1 that measures the correlation between two quantitative variables.

Negative Correlation

A relationship between variables where when one variable moves up or down, the other variable will move in the opposite direction.

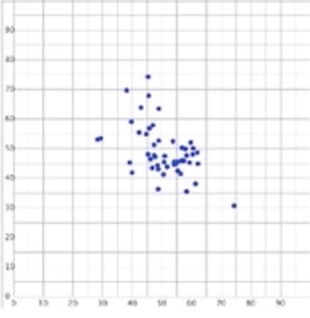
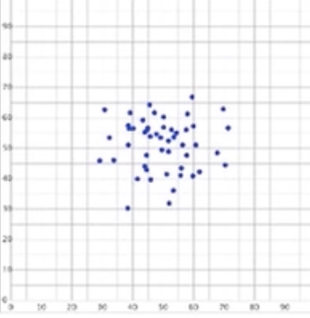
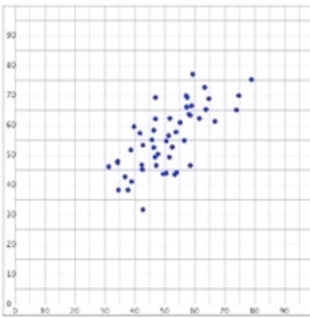
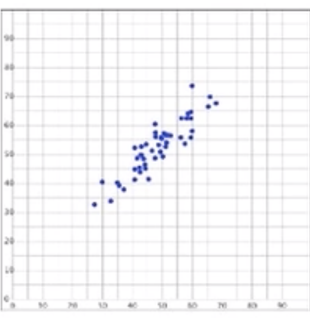
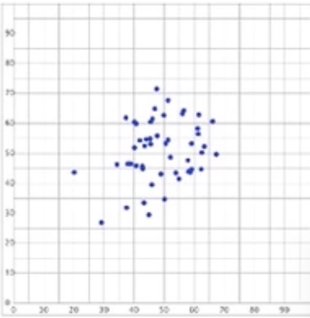
Positive Correlation

A relationship between variables where when one variable moves up or down, the other variable will move in the same direction.

2. Correlation Coefficients and Scatter Plots

Let's explore some scatter plots with different correlation coefficients.

Graph	Correlation	Explanation
	$r = -0.99$	The data points are in a negative direction, so the correlation is a negative number. It is also nearly linear, so its correlation is negative 0.99, which is very close to negative 1. This graph shows a very strong, negative association.

	$r = -0.5$	The data points are in a negative direction, so the correlation is a negative number. However, the data is fairly spread out, so the strength is not terribly strong. This graph shows a weak to moderate, negative association.
	$r = 0$	The data points have a cloudy association, so it is neither positive nor negative. There is also no linear association between the two variables, so the correlation is zero.
	$r = 0.7$	The data points show an upward association, so the correlation is a positive number. Although it is linear, the data points are not very clustered, but still close enough to show a fairly moderate to strong association.
	$r = 0.9$	The data points are in a positive direction, so the correlation is a positive number. It is also very linear, so its correlation will be closer to 1. This graph shows a strong, positive association.
	$r = 0.3$	Even though the points are spread out, a positive association is visible. However, since they are not closely clustered, the data points will show a weaker strength.

The correlation only relates the linear relationship between two *quantitative* variables. As a caution, you're going to hear the word correlation thrown around a lot in everyday speech; however, there are often very common errors made when comparing two different variables. It is always important to make sure the two variables being measured are quantitative.



THINK ABOUT IT

Can you spot the errors in the following statements?

"There is a strong correlation between Type 2 diabetes, physical inactivity, and obesity."

+

Although it's possible that they are related, you can't use the word correlation. The first error here is that three variables are being compared, and correlation only compares two variables. Also, Type 2 diabetes is categorical--either you have it, or you don't. Physical inactivity could be quantitative, but it's not obviously quantitative. Obesity is certainly categorical.

"There is a strong correlation between IQ and religious affiliation."

+

Although IQ is quantitative, religious affiliation is categorical. You can't calculate the correlation between these two data types.



SUMMARY

Correlation measures the strength and direction of a linear relationship between two variables on a scatter plot. Strong associations have correlation coefficients near positive 1 or negative 1. Scatterplots with weak correlation coefficients are values near zero.

Source: THIS TUTORIAL WAS AUTHORED BY JONATHAN OSTERS FOR SOPHIA LEARNING. PLEASE SEE OUR [TERMS OF USE](#).



TERMS TO KNOW

Correlation

The strength and direction of a linear association between two quantitative variables.

Correlation coefficient (r)

The numerical value between -1 and +1 that measures the correlation between two quantitative variables.

Negative Correlation

A relationship between variables where when one variable moves up or down, the other variable will move in the opposite direction.

Positive Correlation

A relationship between variables where when one variable moves up or down, the other variable will move in the same direction.