

Correlation and Causation

by Sophia



WHAT'S COVERED

This tutorial will introduce the connection between correlation and causality. Our discussion breaks down as follows:

1. Correlation and Causation

1. Correlation and Causation

Correlation and causation are not the same thing. However, it's often tempting to say that two well-correlated variables have what we call a "causal" link; that the two variables are causing each other to happen.

⇒ **EXAMPLE** Suppose you have two variables and find that the correlation coefficient is 1, meaning they have a perfect linear correlation and they are strongly associated. However, you cannot say that one variable causes the other variable to happen without doing some other tests and making other assertions.

Correlation is just saying that the two variables or events have a linear association. **Causation** is when one variable causes another variable to occur.

There doesn't always have to be an explanation for the relationship between two events. It's possible that two variables might be very well-correlated, but the correlation is simply a coincidence. Therefore, the best way to prove cause-and-effect is with a controlled experiment where the explanatory variable is administered to one group and withheld from the other.



HINT

Recall that the strength of correlation increases as the correlation coefficient approaches 1 (positive correlation) or -1 (negative correlation). However, even an extraordinarily strong correlation is **NOT** evidence of causation.

If the experiment follows the basic experimental design principles of control, randomization, and replication, the experiment can, in fact, prove a cause-and-effect relationship. It can give the best evidence for causation.



BIG IDEA

Finding a correlation does not mean one variable causes the other causation. In other words, correlation does not imply causation.

IN CONTEXT

Imagine we collect data for monthly ice cream sales and monthly shark attacks around the United States each year. Surprisingly, we find that these two variables are highly correlated. But does this mean that consuming ice cream somehow causes shark attacks? Not quite!

The more likely explanation lies in the weather. During warmer months, people tend to consume more ice cream because it's refreshing. Simultaneously, when it's warm outside, more people venture into the ocean for a swim. And guess what? That's when shark attacks are more likely to occur. So, the correlation between ice cream sales and shark attacks is likely due to a third variable, temperature.

While ice cream sales and shark attacks may dance together in the data, one doesn't directly lead to the other. Instead, it's the shared influence of warm weather that connects them.

If you do find a correlation, there are a variety of explanations for why we cannot say there is causation. In the next lesson you will explore a few of these reasons.



TRY IT

See if you can identify the likely causative link between these correlated variables.

Reading ability in children is strongly correlated to shoe size.

Are kids with big feet better readers?



Nope, it's more likely that older children (who tend to have larger feet) are more advanced in their reading skills due to their age.

The number of firefighters per square mile is correlated with the numbers of fires reported per square mile.

Are firefighters causing fires?



Nope, it's more likely that this correlation is the result of population density. Cities both employ more firefighters and have more opportunity for fires to start.



TERMS TO KNOW

Correlation

A statistic which measures the strength and direction of the linear association between two quantitative variables.

Causation/Cause-and-Effect

A phenomenon whereby an increase in one variable directly leads to an increase or decrease in another variable.



SUMMARY

Sometimes two variables will be related because one causes the other, whereas other times they will be well-correlated, but the association isn't what we call "causal"; this is the difference between correlation and causation.

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



TERMS TO KNOW

Causation/Cause-and-Effect

A phenomenon whereby an increase in one variable directly leads to an increase or decrease in another variable.

Correlation

A statistic which measures the strength and direction of the linear association between two quantitative variables.