

# Lurking Variables

by Sophia Tutorial



## WHAT'S COVERED

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

1. Correlation and Causation
  - a. Lurking Variables
  - b. Reversed Association

## 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 actually 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.

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

Correlation does NOT imply causation!

If you do find a strong correlation, there are a variety of explanations for why we cannot say there is causation:

- There could be something called a "lurking variable" behind the scene that causes an increase or decrease in one or both of them.
- It could simply be that you got the association reversed.



## 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.

#### 1a. Lurking Variables

One reason that we cannot say there is causation with a strong correlation could be a lurking variable. This other variable could be confusing the relationship between the explanatory variable and the response variable.

➞ **EXAMPLE** In many families where parents left the light on in their infant's room as they slept, the infant developed nearsightedness. This is an actual studied scenario, where researchers noticed that there was a positive relationship between sleeping with the light on and having nearsightedness. Therefore, researchers concluded that sleeping with the light on might cause nearsightedness.

Is this conclusion correct?

Upon follow-up studies, this conclusion was shown to be incorrect. The nearsightedness of the children was genetic and was therefore caused by their parents' nearsightedness, not by sleeping in a room with the light on. In fact, the parents' nearsightedness caused them to leave the light on in the child's room so that the *parents* could see.

Therefore, the nearsightedness of the child and the light being left on were both due to the lurking variable of their parents' nearsightedness. It wasn't the light that caused the child's nearsightedness.

➞ **EXAMPLE** As ice cream sales increase, so do the number of drowning deaths. Suppose you come up with this conclusion: "Eating ice cream causes drowning."

So, should you not go swimming after eating ice cream because it's dangerous for you? Well, not really.

Both the variables of ice cream sales and drownings just happen to increase with higher temperatures:

- As the summer months go on, more people consume ice cream because it's warmer and they want to cool off.
- They also want to cool off by going to the beach and the pools in the summer.
- A higher volume of people attending those beaches and pools will sadly cause the number of people that drown to go up, as well.

Just as in the case of the nearsightedness and sleeping with the light on, there's a lurking variable behind the scenes causing the increase in both ice cream sales and drowning. It's not the increase in ice cream sales that causes the drowning, nor does the drowning cause an increase in ice cream sales. They're both increased by the higher temperatures.

### 1b. Reversed Association

Another reason that we cannot say there is causation with a strong correlation could be that the association is reversed. If we don't know the direction of the cause-and-effect of two variables, we cannot say that it is a causal relationship, only that they are strongly correlated.

➡ **EXAMPLE** As the number of firefighters at a fire increases, so does the damage the fire causes. Suppose you come up with this conclusion: "Sending firefighters is counterproductive because they only increase the size of the fire."

This is obviously a ludicrous conclusion to draw. In fact, the true association is just the other way around. The association is reversed. It is cause-and-effect relationship; however, it is a severe fire that causes the firefighters to arrive, not the other way around.



### 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. In many cases, there's a lurking variable--something behind the scenes that's causing an increase or decrease in both variables, or maybe a decrease in one and an increase in the other. Finally, sometimes there appears to be a relationship between two variables, but it is only a coincidence. Thus, the most valid way to prove causation is with a controlled, randomized experiment. However, strong evidence for causation can be made with an observational study.

Good luck!



### 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.