What is the definition and example of the Third Variable Problem?

Authored by

stats writer

April 24, 2024

The Third Variable Problem refers to a common issue in research where a correlation between two variables may be misleadingly attributed to a third, unmeasured variable. This means that the relationship between the two variables is not actually causal, but rather influenced by a separate variable. For example, a study may find a correlation between ice cream sales and crime rates, but the true cause is likely the warmer weather that increases both ice cream sales and crime rates. It is important for researchers to consider and control for potential third variables in order to accurately determine causal relationships between variables.

Third Variable Problem: Definition & Example

In statistics, a third variable problem occurs when an observed correlation between two variables can actually be explained by a third variable that hasn't been accounted for.

When this third variable is not taken into account, the correlation between the two variables under study can be misleading and even confusing.

This tutorial provides several examples of third variable problems in different settings.

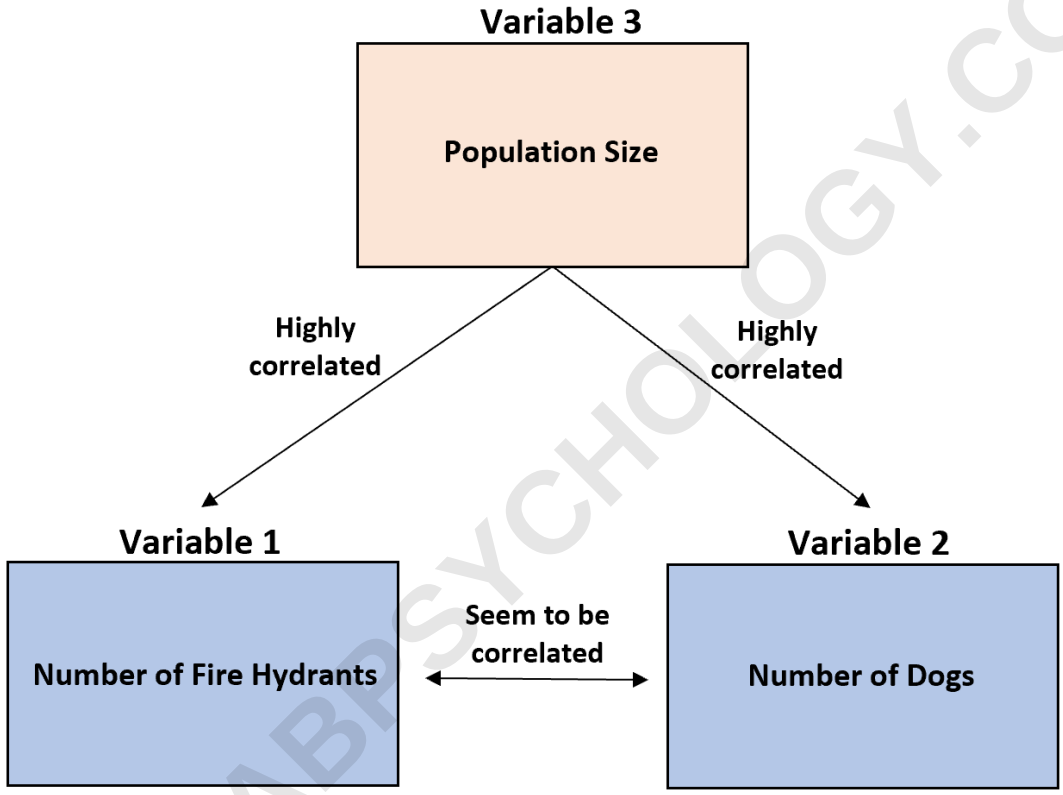
Example 1: Dogs & Fire Hydrants

A researcher observes that cities with more fire hydrants tend to also have more dogs.

However, these two variables are only correlated because they both have a high correlation with a third

variable: population size.

Larger cities tend to have both more fire hydrants *and* more dogs. Conversely, smaller cities tend to have fewer fire hydrants *and* fewer dogs.



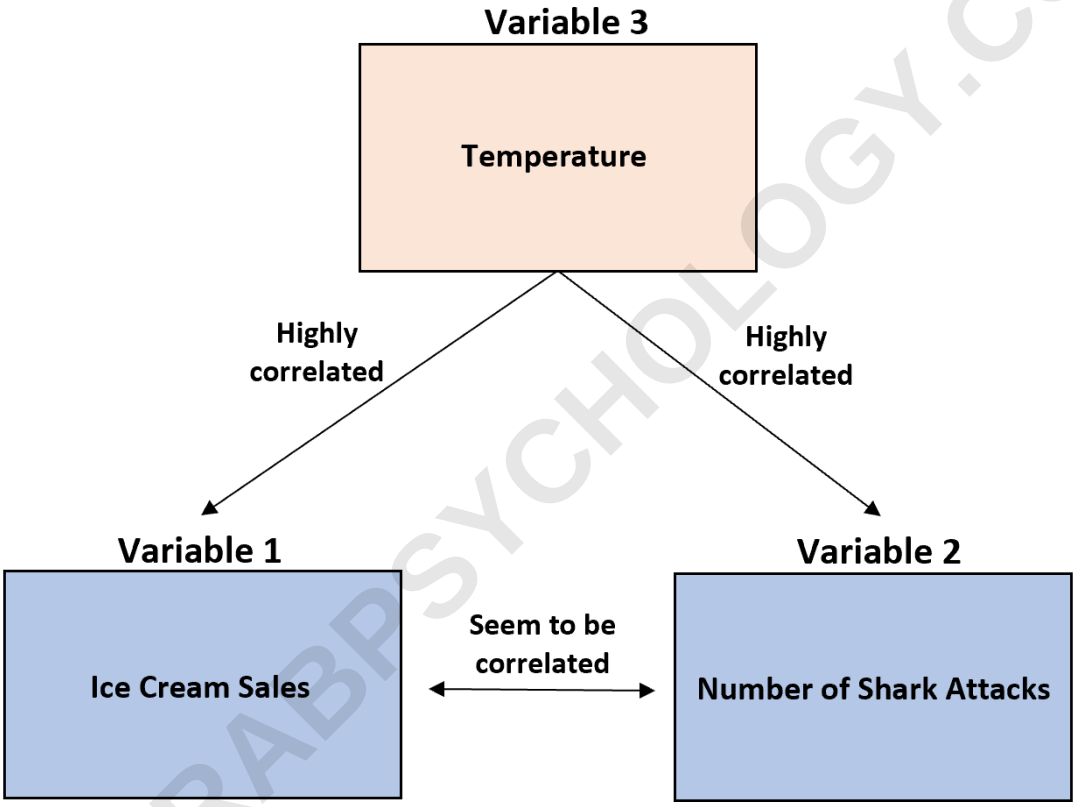
Example 2: Ice Cream Sales & Shark Attacks

A researcher finds that ice cream sales and shark attacks are highly positively correlated.

However, these two variables are only correlated because they both have a high correlation with a third

variable: temperature.

When it's warmer out, more people buy ice cream and more people swim in the ocean which explains why the values for both ice cream sales and shark attacks tend to increase during the same times of the year.



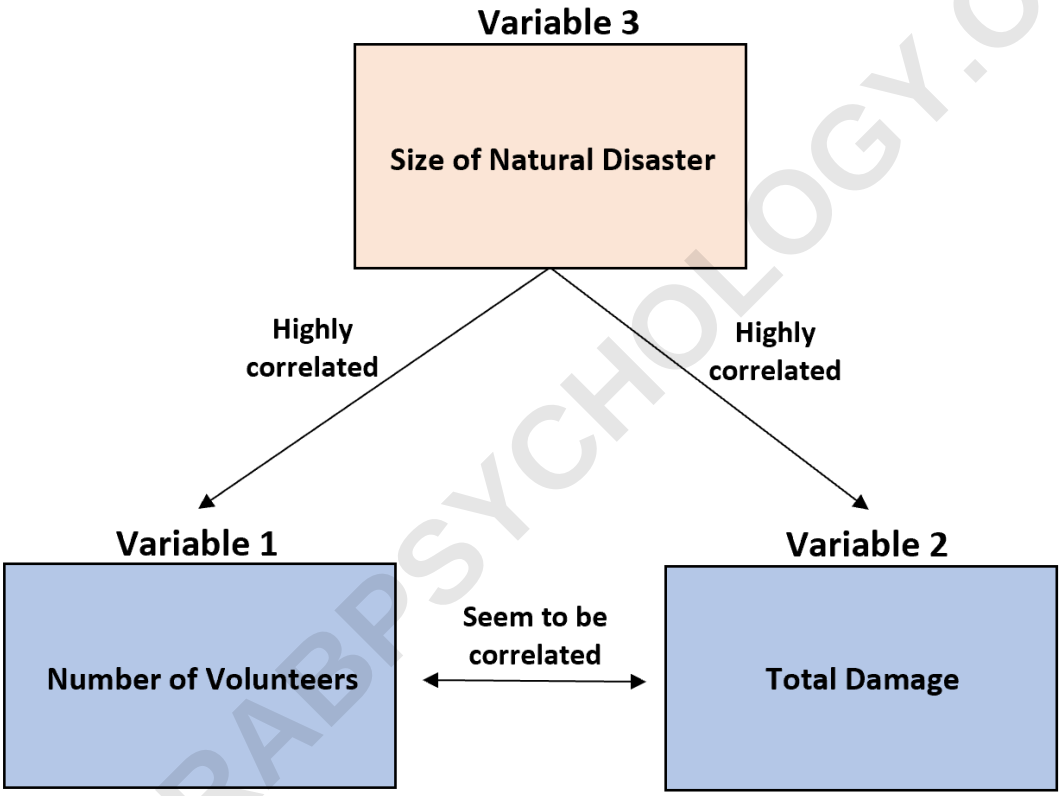
Example 3: Volunteers & Natural Disasters

A study finds that the more volunteers that show up after a natural disaster, the greater the damage.

However, these two variables are only correlated

because they both have a high correlation with a third variable: size of the natural disaster.

Larger natural disasters are highly correlated with more damage done as well as an increase in the number of volunteers.



Related Articles

What is a Confounding Variable?