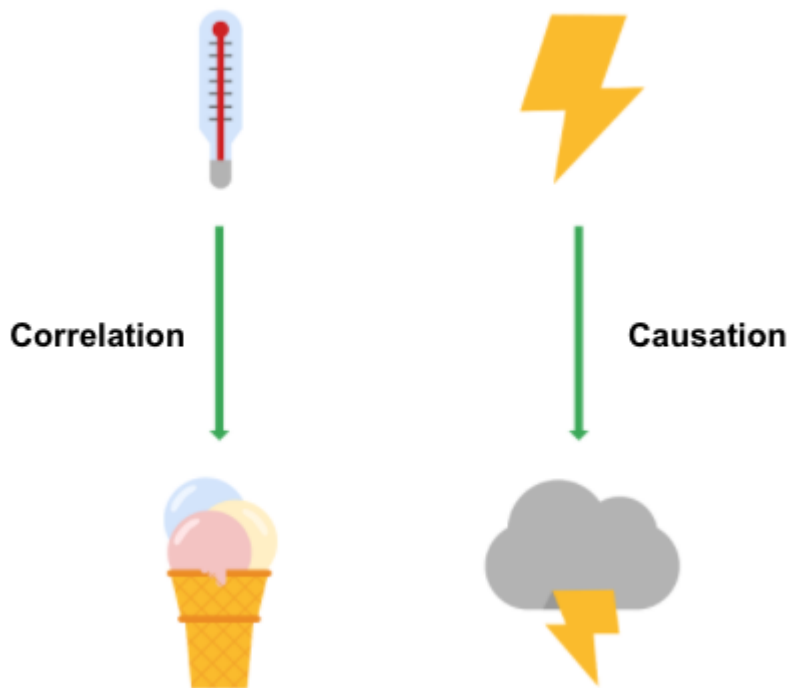


Correlation and causation

In this reading, you will examine correlation and causation in more detail. Let's review the definitions of these terms:

- **Correlation** in statistics is the measure of the degree to which two variables move in relationship to each other. An example of correlation is the idea that "As the temperature goes up, ice cream sales also go up." It is important to remember that correlation doesn't mean that one event causes another. But, it does indicate that they have a pattern with or a relationship to each other. If one variable goes up and the other variable also goes up, it is a positive correlation. If one variable goes up and the other variable goes down, it is a negative or inverse correlation. If one variable goes up and the other variable stays about the same, there is no correlation.
- **Causation** refers to the idea that an event leads to a specific outcome. For example, when lightning strikes, we hear the thunder (sound wave) caused by the air heating and cooling from the lightning strike. Lightning causes thunder.



Why is differentiating between correlation and causation important?

When you make conclusions from data analysis, you need to make sure that you don't assume a causal relationship between elements of your data when there is only a correlation. When your data shows that outdoor temperature and ice cream consumption both go up at the same time, it might be tempting to conclude that hot weather **causes** people to eat ice cream. But, a closer examination of the data would reveal that every change in temperature doesn't lead to a change in ice cream purchases. In addition, there might have been a sale on ice cream at the same time that the data was collected, which might not have been considered in your analysis.

Knowing the difference between correlation and causation is important when you make conclusions from your data since the stakes could be high. The next two examples illustrate the high stakes to health and human services.

Cause of disease

For example, pellagra is a disease with symptoms of dizziness, sores, vomiting, and diarrhea. In the early 1900s, people thought that the disease was caused by unsanitary living conditions. Most people who got pellagra also lived in unsanitary environments. But, a closer examination of the data showed that pellagra was the result of a lack of niacin (Vitamin B3). Unsanitary conditions were related to pellagra because most people who couldn't afford to purchase niacin-rich foods also couldn't afford to live in more sanitary conditions. But, dirty living conditions turned out to be a correlation only.

Distribution of aid

Here is another example. Suppose you are working for a government agency that provides food stamps. You noticed from the agency's Google Analytics that people who qualify for food stamps are browsing the official website, but they are leaving the site without signing up for benefits. You think that the people visiting the site are leaving because they aren't finding the information they need to sign up for food stamps. Google Analytics can help you find clues (correlations), like the same people coming back many times or how quickly people leave the page. One of those correlations might lead you to the actual cause, but you will need to collect additional data, like in a survey, to know exactly why people coming to the site aren't signing up for food stamps. Only then can you figure out how to increase the sign-up rate.

Key takeaways

In your data analysis, remember to:

- Critically analyze any correlations that you find
- Examine the data's context to determine if a causation makes sense (and can be supported by all of the data)
- Understand the limitations of the tools that you use for analysis

Further information

You can explore the following article and training for more information about correlation and causation:

- **Correlation is not causation**: This article describes the impact to a business when correlation and causation are confused.
- **Correlation and causation (Khan Academy lesson)**: This lesson describes correlation and causation along with a working example. Follow the examples of the analysis and notice if there is a positive correlation between frostbite and sledding accidents.