Causation is functional and naturally would reflect correlation, but it doesn’t indicate causation because it doesn’t go beyond studying covariation.

**Causation:** X is the cause of change in Y i.e, the change of Y is the effect of change in X.

***NOTE:***

–  If X and Y are correlated then X and Y may or may not have a casual relationship.

– If X and Y have a causal relationship then X and Y must be correlated.

The degree of correlation between two or more variables can be determined using correlation. However, it does not consider the cause-and-effect relationship between variables. If two variables are correlated, it could be for any of the following reasons:

**1. Third-Party Influence:**

The influence of a third party can result in a high degree of correlation between the two variables. This analysis does not take into account third-party influence. ***For example,*** the correlation between the yield per acre of grain and jute can be of a high degree because both are linked to the amount of rainfall. However, in reality, both these variables do not have any effect on each other.

**2. Mutual Dependence (Cause and Effect):**

It may be challenging to determine which is the cause, and which is the effect when two variables indicate a high degree of correlation. It is so because they may be having an impact on one another. ***For example,***when there is an increase in the price of a commodity, it increases its demand. Here, the price is the cause, and demand is the effect. However, there is a possibility that the price of the commodity will rise due to increased demand (population growth or other factors). In that case, increased demand is the cause, and the price is the effect.

**3. Pure Chance:**

It is possible that the correlation between the two variables was obtained by random chance or coincidence alone. This correlation is also known as **spurious**. Therefore, it is crucial to determine whether there is a possibility of a relationship between the variables under analysis. ***For example,*** even if there is no relationship between the two variables (between the income of people in a society and their clothes size), one may see a strong correlation between them.

So, it can be said that correlation provides only a quantitative measure and does not indicates cause and effect relationship between the variables.

” We study causation because we need to make sense of data, to guide actions and policies, and to learn from our success and failures. We need to estimate the effect of smoking on lung cancer, of education on salaries, of carbon emissions on the climate. Most ambitiously, we also need to understand how and why causes influence their effects, which is not less valuable. For example, knowing whether malaria is transmitted by mosquitoes or “mal-air,” as many believed in the past, tells us whether we should pack mosquito nets or breathing masks on our next trip to the swamps. Less obvious is the answer to the question, “why study causation as a separate topic, distinct from the traditional statistical curriculum?” What can the concept of “causation,” considered on its own, tell us about the world that tried-and-true statistical methods can’t? Quite a lot, as it turns out. When approached rigorously, causation is not merely an aspect of statistics; it is an addition to statistics, an enrichment that allows statistics to uncover workings of the world that traditional methods alone cannot. For example, and this might come as a surprise to many, none of the problems mentioned above can be articulated in the standard language of statistic

Why Establishing Correlation vs Causation is Important

Correlation only indicates that two variables move together, but it doesn’t tell us if one causes changes in the other. Relying solely on correlation can lead to misguided conclusions and ineffective or even harmful actions. Establishing causality ensures that we’re targeting the root cause of an issue rather than just an associated symptom.

For instance, a study published in the journal [Language Sciences](https://www.sciencedirect.com/science/article/abs/pii/S038800011400151X) found a correlation between individuals who use taboo words (swear words) and higher levels of verbal intelligence. However, it’s essential to approach such findings with nuance. The correlation doesn’t suggest that swearing enhances intelligence. Instead, it might indicate that individuals with a richer vocabulary, encompassing both standard and taboo words, have a more extensive linguistic repertoire to express themselves. Misunderstanding this correlation could lead to the mistaken belief that increasing one’s use of swear words would boost intelligence, which is not what the study implies.

Alternatively, suppose you unknowingly find a spurious correlation between vitamins and improved health outcomes. Believing that the vitamins cause those improvements when it’s merely correlation leads to poor decision-making. After all, if the vitamins don’t *cause* health gains, then consuming more vitamins won’t produce better outcomes despite the correlation.

Scientists have found that people who regularly take vitamins have many pre-existing health habits and conditions that differ from non-vitamin consumers. Those differences are the likely causes for the improved health outcomes rather than the vitamins themselves. Read my posts about [Observational Studies](https://statisticsbyjim.com/basics/observational-studies/) to see the long list of differences in the example.

These examples underscore the critical importance of distinguishing between correlation vs causation in decision-making.

How to Identify Causal vs. Correlational Relationships

Establishing correlation vs causation is often misunderstood. While correlation can provide hints about potential relationships between variables, it doesn’t prove that one variable causes another to change. That’s an entirely different matter. They might not be causally linked at all. Unfortunately, spurious correlations occur frequently, and there’s no statistical test for detecting them!

So, how do you distinguish between correlation vs causation?

To truly establish causality, researchers need specially designed experiments—randomized controlled trials (RCTs). RCTs randomly assign participants to either a treatment or control group. This random assignment helps ensure all groups start the same except for the treatment. If the outcomes differ at the end, analysts can attribute them to the treatment with high confidence. Learn more about [Randomized Controlled Trials](https://statisticsbyjim.com/basics/randomized-controlled-trial-rct/) and [Random Assignment in Experiments](https://statisticsbyjim.com/basics/random-assignment-experiments/).

Conversely, [Correlational Studies](https://statisticsbyjim.com/basics/correlational-study/) are suited for finding relationships quickly and inexpensively in preliminary studies but they are not suitable for establishing causality.

Sir Austin Bradford Hill proposed a set of nine criteria to help determine if a relationship is genuinely causal rather than merely correlational. These criteria are an exercise in critical thought. They prompt you to think about causation by highlighting vital properties to consider and how to apply your subject-area knowledge. The objective is to fulfill as many criteria as possible. While no single criterion is adequate, it’s usually impossible to meet all of them. For a deeper dive, read my post about [Causation in Statistics: Hill’s Criteria](https://statisticsbyjim.com/basics/causation/).

Regardless of the method used, it’s crucial to approach the question of correlation vs causation with caution, ensuring that you base your conclusions on solid evidence, sound methodology, and critical thinking.

**Key Principles of Causation**

**1. Temporal Precedence**: The cause must happen before the effect. This helps us understand which one comes first and shows us the direction of cause and effect.

**2. Covariation**: Changes in the cause should always be followed by changes in the effect. Methods like long-term studies can prove this connection over time.

**3. Controlling for Confounding Factors**: It’s important to find other factors that could affect the relationship between variables. Methods like statistical controls and random selection can reduce the impact of these factors.

**Establishing Causation**

To establish causation, researchers often use experimental designs. They change the independent variable and observe changes in the dependent variable. Key methods include:

**Randomized Controlled Trials (RCTs)**: These experiments randomly assign participants to different groups. One variable is changed, and other variables constant are kept constant.

**Longitudinal Studies**: These studies track variables for a long period of time. It examines how changes in one variable (independent variable) relate to changes in another variable (dependent variable).

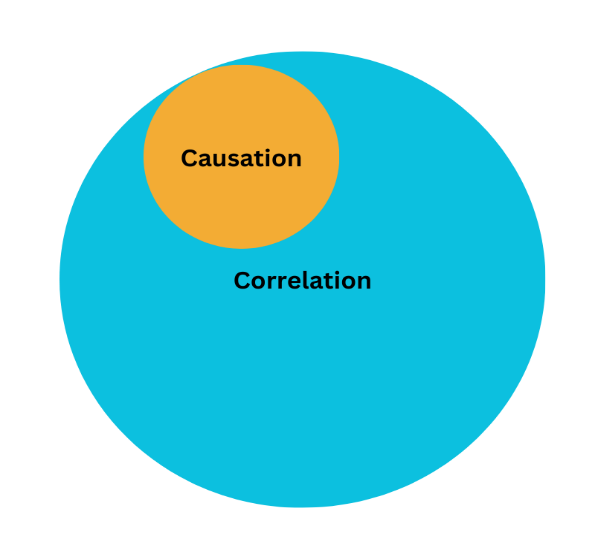
**Challenges and Considerations**

**Ethical Constraints**: Some experiments that establish causation may be unethical to conduct.

**Complexity**: It is difficult to establish causation due to the presence of multiple variables and interactions among them.

**Why Correlation Doesn’t Imply Causation**

Correlation does not always mean causation. Causation usually mean correlation. The Venn diagram shows the overlap between them.



Let’s explore why spurious correlations happen.

**Wrapping Up**

Correlation means two things happen together. On the other hand, causation means one thing makes another thing happen. Correlation doesn’t necessarily imply causation. Just because two things happen together doesn’t mean that one causes the other. It is essential to understand this distinction for accurate data analysis.

[Correlation vs Causation: Understanding the Differences - Statistics By Jim](https://statisticsbyjim.com/basics/correlation-vs-causation/)