

Correlation Does Not Prove Causation

A correlation coefficient describes a pattern between two variables. It cannot, by itself, tell you which of several different explanations produced that pattern.



StatReason

Correlation Does Not Prove Causation

Exported from statreason.com. Deterministic, browser-local content; not professional statistical or research advice.

What this answers

This page answers "if X and Y are correlated, does X cause Y?" A correlation coefficient measures only the strength and direction of a statistical association. At least four distinct explanations, only one of which is a direct causal effect, can produce the same observed correlation.

Four explanations for the same correlation

X could genuinely cause Y. Alternatively, the direction could be reversed: Y could cause X, which is easy to overlook when the data are collected at a single point in time. A third variable could affect both X and Y independently, creating a correlation between them with no direct link at all, a pattern called confounding. Finally, a selected or non-random sample can create an apparent association that would not exist in the full population, and pure chance can produce an unusual pattern in any single sample, particularly a small one.

What actually supports a causal claim

Supporting a causal claim generally requires a randomized experiment, where assignment to conditions is controlled and not related to any other characteristic of the participants, or a well-justified quasi-experimental design that addresses confounding and reverse causation directly, such as a natural experiment with a credible comparison group. Time order, a plausible mechanism, and ruling out major confounders all strengthen a causal argument, but none of them come from the correlation coefficient itself.

Worked example

Ice cream sales and drowning incidents are positively correlated across months. Neither causes the other directly; a third variable, warm weather, increases both the number of people swimming and the number of people buying ice cream. Treating this correlation as evidence that ice cream causes drownings would be a confounding error, the kind this page is meant to help you avoid.

Assumption audit

Calculated from your data: the correlation coefficient, sample size, and confidence interval, once you run the linked calculator on your own two variables.

Evidence to review: whether a scatter plot suggests a linear or a curved pattern, and whether the timing of measurement makes reverse causation plausible.

You must verify: the study design behind your data, specifically whether conditions were randomly assigned, since no correlation coefficient can substitute for that design fact.

Source

This explanation follows the standard causal-inference caveats in the NIST/SEMATECH e-Handbook of Statistical Methods and the shared statistical reasoning contract every StatReason engine is built against.

Limitations

This page explains why correlation alone cannot establish causation; it does not cover the specific statistical methods, such as randomized controlled trials or instrumental variable analysis, used to support causal claims in practice.

Next action: measure the association with the [Pearson Correlation Calculator](#) or [Spearman Correlation Calculator](#), then read [Pearson vs Spearman Correlation](#) to choose between them.