Correlation vs Causation

Correlation means that two things tend to move together, such as study time and exam marks. Causation means that a change in one thing actually produces a change in the other. The two are often mixed up, and mixing them up can lead to costly decisions. We follow Dina, a marketing analyst at a retailer, who has to explain what her data really shows.

1. What Correlation Shows

Correlation describes how two measures move together. In a positive correlation they rise together, in a negative correlation one rises while the other falls, and with no correlation there is no pattern. Statisticians often summarise the strength with a number between minus one and plus one, where values near the ends mean a strong link and values near zero mean a weak one. This number describes a pattern in the data, and it does not explain why the pattern exists.

Example: A café sells less hot soup whenever the outside temperature rises: as one goes up, the other goes down, which is a negative correlation. The pattern is clear, and it is easy to see why.

At work: Dina finds that weeks with more advertising spend also have higher sales, with a correlation of about zero point eight. This is a strong positive link, but she does not yet know why the two move together.

2. Why It Is Not Proof

There are three common reasons why a correlation may not mean cause. The first is coincidence: with enough data, some unrelated things will move together by chance. The second is a hidden third factor, called a confounder, which influences both things. The third is reverse direction, where the effect is actually causing the supposed cause. A strong correlation can have any of these explanations, so it is only a starting point for questions.

Example: Ice cream sales and drownings both rise in summer, but ice cream does not cause drowning. Hot weather is the hidden third factor that increases both.

At work: Dina notices that customers who open many emails also spend more, so she wonders whether the emails cause the extra spending. The hidden third factor may be loyalty: loyal customers both open emails and buy more. The correlation alone cannot tell the two explanations apart.

This document is a property of Asaseeyat Academy - © Copy Right Protected.

3. How to Test for Cause

The strongest way to test for cause is an experiment. We split people at random into two groups, give the change to one group and compare the results. Random assignment makes the groups alike in everything else, so a difference is more likely to come from the change. When an experiment is not possible, careful comparisons, timing, a believable mechanism and repeated results give weaker but useful evidence.

Example: To test whether a new fertiliser works, a gardener treats half of the plants, chosen at random, and leaves the other half untreated. If the treated half grows better under the same conditions, the fertiliser is a likely cause.

At work: Dina runs an A/B test: a random half of the customers receive the email and the other half do not. The group with the email buys more than the group without it. Because the groups were chosen at random, she can say that the email probably helped.

4. Report with Care

Careful wording protects decisions. We can say that two things are associated or linked, and we should keep words such as causes or leads to for cases with strong evidence. A good report also names the possible hidden factors and the limits of the data. Finally, it suggests the next step, such as a test, before money or policy depends on the result.

Example: A headline says that people who eat breakfast do better at school. A careful version says that breakfast and school results are linked, and that family habits may play a role.

At work: Dina writes in her report that email openers spend more, and that this is a link, not proof. She lists loyalty as a possible hidden factor and shows the test result. The manager then funds the email plan with more confidence.

Conclusion

In summary, correlation shows that two things move together, while causation means that one produces the other. Four steps put this into practice: see the pattern, question it, test for cause and report with care. Used this way, data leads to sound decisions rather than costly guesses.

Sources

• Hill, A. B. (1965). The Environment and Disease: Association or Causation? Proceedings of the Royal Society of Medicine, 58(5), 295-300 – https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1898525/

• Correlation does not imply causation - overview of coincidence, confounders and reverse causation (Wikipedia) – https://en.wikipedia.org/wiki/Correlation_does_not_imply_causation

• Confounding - how a hidden third factor can create a false link (Wikipedia) – https://en.wikipedia.org/wiki/Confounding

This document is a property of Asaseeyat Academy - © Copy Right Protected.