

Korrelation und Kausalität

Korrelation und Kausalität sind zwei zentrale Konzepte aus der Statistik, die oft verwechselt werden. Um keine voreiligen Schlussfolgerungen zu ziehen, wenn Daten statistische Muster aufzeigen, ist es wichtig, ihren Unterschied zu verstehen.

Korrelation ist eine empirische Beobachtung, während Kausalität einen Ursache-Wirkungs-Zusammenhang beschreibt.

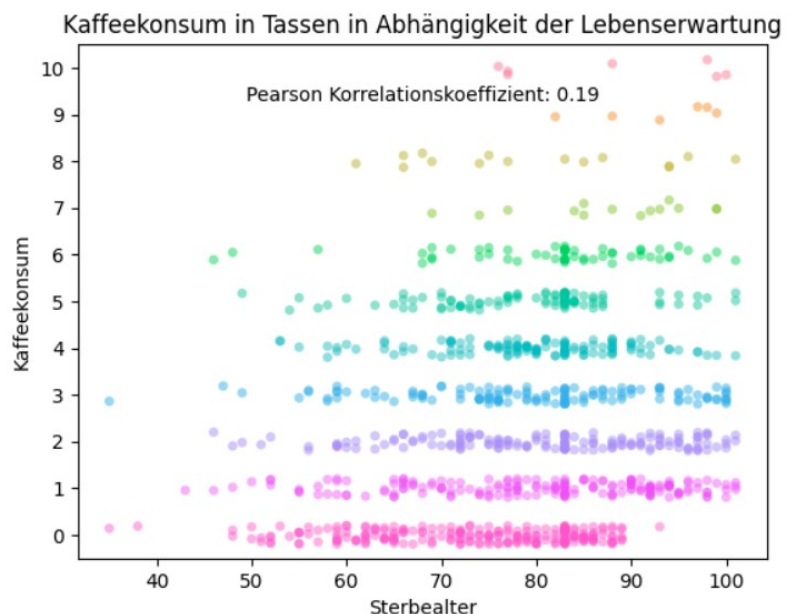
Nehmen wir an, wir hätten aus einer großen Bevölkerungsstudie Daten über den Kaffeekonsum und die Lebensdauer von eintausend Personen.

Personen-ID	Kaffeekonsum pro Tag in Tassen	Lebensdauer in Jahren
1	3	84
2	6	90
3	1	79
4	0	45
...

Korrelation beschreibt die statistische Beziehung zwischen mehreren Variablen (Datenreihen). Wenn zwei Variablen korreliert sind, bedeutet das, dass eine Veränderung in einer Variable mit einer Veränderung in der anderen Variable einhergeht. Gibt es also ein Muster zwischen den Variablen „Kaffeekonsum“ und „Lebensdauer“? Wenn beide Variablen gemeinsam steigen oder fallen, sind sie positiv korreliert, verhalten sie sich entgegengesetzt, so spricht man von negativer Korrelation. In unserem Beispiel gibt es eine positive Korrelation zwischen dem Kaffeekonsum und der Lebenserwartung, da Personen, die mehr Kaffee konsumieren, tendenziell länger leben.

Die Korrelation zwischen Variablen wird mithilfe von Korrelationskoeffizienten berechnet – das sind Maßzahlen, die die Stärke und Richtung der Korrelation angeben. Der Pearson-Korrelationskoeffizient liegt im Bereich von -1 bis +1, wobei positive Werte auf eine positive lineare Beziehung und negative Werte auf eine negative lineare Beziehung hinweisen. Ein Wert von 0 deutet auf keine lineare Beziehung hin.

Korrelation betrachtet also, wie sich Variablen gemeinsam verhalten, ohne direkt auf eine Ursache-Wirkungs-Beziehung einzugehen.



Kausalität hingegen bedeutet, dass eine Variable tatsächlich einen Einfluss auf eine andere Variable hat. Wenn der Kaffeekonsum die Veränderung der Lebenserwartung verursacht, besteht eine kausale Beziehung zwischen ihnen. Kausalität erfordert eine umfassendere Beweisführung, um eine direkte Verbindung zwischen den Variablen zu etablieren. Um Kausalität zu behaupten, müssten wir nicht nur zeigen, dass ein höherer Kaffeekonsum mit einer längeren Lebenserwartung einhergeht, sondern wir müssten auch andere potenzielle Einflussfaktoren ausschließen und eine plausible Erklärung für die Ursache-Wirkungs-Beziehung liefern. Hierfür werden statistische Methoden wie Regressionsanalysen, randomisierte Kontrollstudien oder Pfadanalysen eingesetzt. Diese Methoden dienen dazu, potenzielle Störfaktoren zu kontrollieren, alternative Erklärungen auszuschließen und die Plausibilität einer kausalen Beziehung zwischen den Variablen zu stärken.

Da mit Korrelation manchmal fälschlicherweise Kausalität impliziert wird, betonen Artikel über Datenanalyse zu Recht, dass Korrelation nicht zwangsläufig auf Kausalität hindeutet. Nur weil zwei Variablen miteinander korreliert sind, bedeutet das nicht, dass eine Variable die Ursache für die Veränderung der anderen ist. Es könnte auch andere Faktoren geben, die die Beziehung zwischen den Variablen beeinflussen – im Fall unseres Beispiels zum Kaffeekonsum und der Lebenserwartung könnten beispielsweise genetische Veranlagungen, Lebensstilfaktoren oder sozioökonomische Bedingungen eine Rolle spielen. Bei der Kommunikation von Analyseergebnissen ist es daher wichtig, vorsichtig zu formulieren.

Takeaway:

❌ Sage nicht: Ein höherer Kaffeekonsum führt zu einer längeren Lebenserwartung.

✅ Sage stattdessen: Es gibt eine positive Korrelation zwischen dem Kaffeekonsum und der Lebenserwartung, da Personen, die mehr Kaffee konsumieren, tendenziell eine höhere Lebenserwartung aufweisen. Dennoch können andere Faktoren die Beziehung beeinflussen, und es bedarf weiterer Untersuchungen, um eine kausale Verbindung festzustellen.

Für weitere Informationen über Fehlschlüsse siehe [Wikipedia: Cum hoc ergo propter hoc](#)