

Definitionen

Korrelation/-Korrelationsbeziehung Beziehung zweier Variablen/Eigenschaften.

Positive Korrelation Steigt eine Variable, steigt auch die andere.

Negative Korrelation Steigt eine Variable, sinkt die andere.

Kausalität/Kausalitätsbeziehung Beziehung zweier Variablen, wobei die Veränderung eine Variable, bedingt dass die andere sich verändert.

Störvariable/Störfaktor (eng. confounding variable) Eine dritte, unbekannt Variable, welche Auswirkung auf die bekannten Variablen hat.

Achtung: Positive/Negative Korrelation bedeutet NICHT, dass eines das andere verursacht, lediglich dass sie Positiv oder Negativ Korrelieren! z. B.: Verkaufte Eiscreme und Zahl der ertrunkenen Personen korrelieren Positiv, bedingen sich jedoch nicht.

Daten Hinterfragen

Gibt es eine Plausible Erklärung? Gibt es eine plausible Erklärung wieso die daten Korrelieren?

Wurden die Kausalitäten belegt? Wenn nur die Korrelation aufgezeigt, jedoch nicht belegt wird, ist das ein schlechtes Zeichen.

Allgemein Skeptisch sein! Daten und vorallem Grafiken nicht einfach akzeptieren, kritische Fragen stellen!

Wissen: Korrelation nicht gleich Kausalität: Auch wenn Variablen eine Korrelationsbeziehung haben, müssen sie keine Kausalitätsbeziehung haben.

Typische Fehler

Korrelation als Beweis Korrelationen sind in erster Linie Hypothesen, welche durch Experimente bewiesen werden müssen.

Zu viele Daten Wenn man mehr Daten beobachtet, finden sich immer leichter spontane Korrelationen die nichts bedeuten.

Zufall Korrelationen können auch durch puren Zufall entstehen, daher immer die Plausibilität hinterfragen.

Unbekannte Störvariable Variablen können den anschein einer Kausalitätsbeziehung haben, jedoch hängen sie lediglich beide von einer dritten, unbekannt ab.

Kausalität Beweisen

Option 1: Experiment Unter Kontrolle der Störvariablen, wird untersucht, ob eine Veränderung der Variable A, eine Veränderung in B auslöst oder vice versa.

Option 2: Beobachtung Wenn kein Experiment durchgeführt werden kann, kann man verschiedene Kausalhypothesen durch Beobachtung überprüfen.

Experimente sind grundsätzlich die bessere Option zur Beweisung der Kausalität, da sie Störvariablen entfernen. Beobachtungen können durch unbekannt Variablen gestört werden.

Nützliche Ressourcen

Spurious Correlations (www.tylervigen.com/spurious-correlations) Webseite, welche eine Liste von Korrelationen zeigt, welche keine Kausalitätsbeziehung haben.

Faktenchecker (Correctiv.org, Saferinternet.at, FactCheck.org) Auf Faktenchecker seiten lassen sich viele Beiträge finden, welche die Korrelation als Kausalität verwechseln.

Artikel zu Korrelation/-Kausalität bei Big Data <https://www.kdnuggets.com/2019/09/risk-ai-big-data.html>