

Das Problem der Kausalität in den Sozialwissenschaften: Die Interpretation von Zusammenhangsmaßen (Korrelationen)

Wenn ein statistisch signifikanter Zusammenhang zwischen zwei Variablen X und Y beobachtet wird (wie es z.B. in Fragebogenstudien häufig der Fall ist), sind prinzipiell vier Kausalinterpretationen möglich:

- 1) X beeinflusst Y kausal;
- 2) Y beeinflusst X kausal;
- 3) X und Y werden von einer dritten oder von weiteren Variablen kausal beeinflusst;
- 4) X und Y beeinflussen sich gegenseitig (kausal).

Der Korrelationskoeffizient liefert keine Information darüber, welche der vier Interpretationen die richtige ist. Eine Korrelation zwischen zwei Variablen ist eine notwendige, aber keine hinreichende Bedingung für Kausalzusammenhänge..

Wenn also zwischen zwei Variablen X und Y eine Korrelation beobachtet wird, darf nicht ohne Weiteres geschlussfolgert werden, dass ein Kausalzusammenhang zwischen X und Y besteht. Korrelationen können nur etwas über die Stärke und die Richtung des Zusammenhangs zeigen (positiv oder negativ), aber nichts über seine Ursache. Signifikante Korrelationen sind kein Beweis für Ursache-Folge-Relationen, weil für einen beobachteten Zusammenhang unendlich viele (theoretische) Kausalmodelle konstruiert werden können. Allerdings können durch nicht signifikante Korrelationen unter bestimmten Bedingungen Kausalzusammenhänge falsifiziert (widerlegt) werden. Wenn z.B. behauptet wird, dass übermäßiger Alkoholkonsum die Lebenserwartung verkürzt (*Modell a*), kann diese Kausalhypothese widerlegt werden, wenn keine Korrelation zwischen Alkoholkonsum (X) und Lebenserwartung (Y) besteht. Aber wenn Lebenserwartung und die Menge des Alkoholkonsums signifikant miteinander zusammenhängen, konnte zum einen das angenommene Kausalmodell nicht widerlegt werden und zugleich sind jedoch auch andere Kausalzusammenhänge möglich. Beispiele hierfür sind die Modelle *b*) bis *f*):

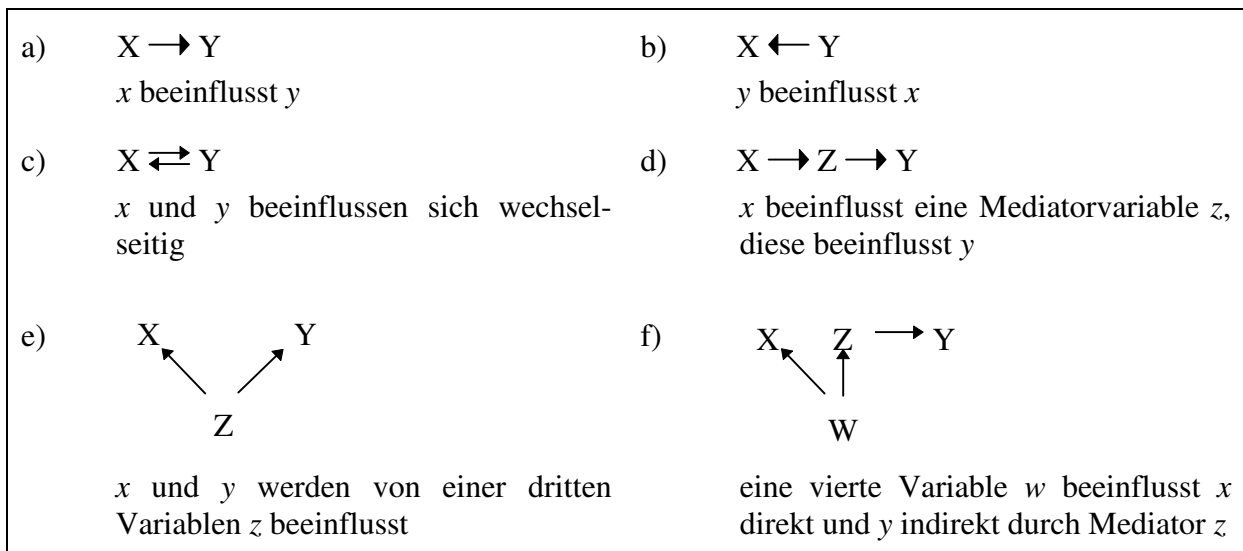
Model b: Eine kurze Lebenserwartung erhöht den Alkoholkonsum.

Model c: Erhöhter Alkoholkonsum verursacht Depression und verkürzt die Lebenserwartung. Das Bewusstsein einer verkürzten Lebenserwartung verstärkt wiederum den Alkoholkonsum.

Model d: Erhöhter Alkoholkonsum führt zu Arbeitsunfähigkeit und Armut (Z). Armut verursacht eine ungesunde Ernährung, wodurch die Lebenserwartung geringer wird.

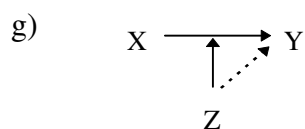
Model e: Eine angeborene Charakterschwäche erhöht sowohl die Anfälligkeit für Krankheiten als für Alkoholkonsum.

Model f: Stress (W) verursacht Alkoholmissbrauch und Rauchen (Z). Die Lebenserwartung wird allein durch Rauchen verkürzt.



Welches dieser Kausalmodelle das plausibelste ist wird nicht durch die Korrelation zwischen X und Y sondern allein durch das Wissen des Forschers bzw. seine Theorie bestimmt. Nur Logik oder konsistente Überlegungen und Schlussfolgerungen können die Basis für die Kausalinterpretation von Korrelationen sein.

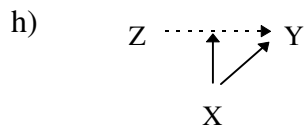
Moderatorvariable (Modell g): Aber selbst wenn keine Korrelation zwischen X und Y gefunden wird, besteht die Möglichkeit, dass eine dritte Variable der Grund dafür ist, dass ein erwarteter Zusammenhang nicht beobachtet wird. Es könnte sein, dass Armut (Z) den Zusammenhang zwischen X und Y derartig moderiert, dass für niedrige Werte von Z (= ein gewisser Wohlstand) kein Zusammenhang und für hohe Werte von Z (= Armut) ein negativer Zusammenhang existiert. In diesem Fall handelt es sich um eine **Interaktion** zwischen X und Z . D.h., der Zusammenhang zwischen X und Y unterscheidet sich in Abhängigkeit der Werte von Z (*Modell g*). Die oben genannte Erwartung (Alkoholkonsum verkürzt die Lebenserwartung) kann nur für niedrige Werte von Z widerlegt werden. Mit anderen Worten, die Zusammenhangshypothese gilt in diesem Fall nur unter bestimmten Bedingungen.



x beeinflusst y in Abhängigkeit der Werte der Moderatorvariable z

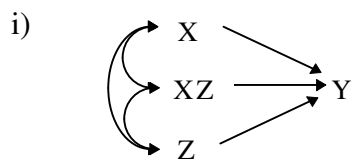
Weil Interaktionseffekte auf unterschiedliche Weise interpretiert werden können, muss hierauf noch genauer eingegangen werden. Das oben gezeigte Modell g stellt die Interaktion von X und Z in Form eines Pfadmodells dar. Weil Interaktionseffekte mathematisch gesehen symmetrisch sind unterscheidet sich Modell g in statistischer Hinsicht nicht von Modell h : Hier wird X als die Moderatorvariable betrachtet, die den Zusammenhang zwischen Z und Y moderiert. Die Größe der Interaktion bzw. der Moderatoreffekte sind in den Modellen g und h exakt gleich. Ob der Forscher die Interaktion gemäß Modell g oder gemäß Modell h inter-

interpretiert (“Armut Z moderiert den Zusammenhang zwischen Alkoholkonsum X und Lebenserwartung Y ” oder “Alkoholkonsum X moderiert den Zusammenhang zwischen Armut Z und Lebenserwartung Y ”) hängt einzig von theoretischen Erwägungen ab.



z beeinflusst y in Abhängigkeit der Werte der Moderatorvariable x

Der Grund für diese zwei Möglichkeiten, die Interaktion von X und Z zu interpretieren (vgl. die Modelle g und h) liegt in der Methode, mit der der Interaktionseffekt berechnet wird. Die Größe des Interaktionseffekts ist nämlich die Größe des Regressionseffekts von Y auf das Produkt $X \cdot Z$ – in dem X und Z das gleiche Gewicht haben –, wobei die im Modell ebenfalls berechneten Effekte von Y auf X und Z auch als Haupteffekte bezeichnet werden (Modell i)¹.



Regressionsmodell mit den Haupteffekten X und Z und einem Interaktionseffekt von X und Z (Produkt $X \cdot Z$).

¹ In den Modellen g , h und i werden die Variablen X und Z als kontinuierlich unterstellt. Aber die Variablen X und Z können auch nominale Variablen sein. In diesem Fall werden ihre Haupteffekte und ihre Interaktion anhand einer Varianzanalyse bestimmt.