

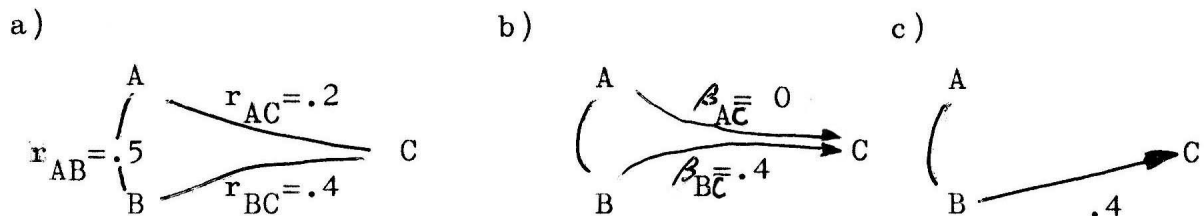
Die Arbeit entstand anlässlich einer Klausurarbeit bei Prof. Dr. P. Heintz (1979).
Erläuterungen dazu siehe Kapitel 6 (Klausurthema am Schluss)
Sie wurde dann für diese Veröffentlichung 1990 überarbeitet.

Die Pfadanalyse als Mittel zur Erforschung kausaler Zusammenhänge in der Soziologie

Richard Grünenfelder

1. Einführung

Die Pfadanalyse ist eine Methode, um lineare Beziehungen zwischen einer Menge von Variablen zu zerlegen und zu interpretieren. Unter anderem wird eine schwache kausale Ordnung vorausgesetzt. Mit dieser Methode ist es möglich die Grösse des direkten und indirekten Einflusses, den jede Variable auf andere Variablen hat, zu bestimmen. Auf diese Art erhalten wir schlussendlich die Pfadkoeffizienten als Masse des direkten Einflusses der jeweiligen Variablen.



Für ein Dreivariablen-system seien die in Abb. 1a angegebenen einfachen Korrelationskoeffizienten ermittelt worden. Wir betrachten C als die abhängige und A und B als die unabhängigen Variablen. Beim Betrachten des Zusammenhangs von A und C, ist A einerseits direkt und andererseits indirekt über B mit C verbunden. Der einfache Korrelationskoeffizient $r_{AC} = 0.2$ suggeriert uns, dass von A auf C ein positiver, direkter Einfluss besteht. Die Methode der Pfadanalyse berücksichtigt nun auch den indirekten Einfluss über B und errechnet den tatsächlich direkten Pfadkoeffizienten $\beta_{AC} = 0$ (vgl. Abb. 1b). Wir können deshalb den Pfeil von A nach C weglassen und erhalten für unser Dreivariablen-system das Pfadschema in Abb. 1c).

2. Die Berechnung der Pfadkoeffizienten

Anhand eines Vier-Variablen-Systems möchte ich kurz auf die Berechnung der Pfadkoeffizienten eingehen.

1) Die vier Variablen sollen in unstandardisierten Werten vorliegen. Wir legen sodann die kausale Folge der Variablen fest und entscheiden uns z.B. für das untenstehende rekursive Kausalmodell. Dabei erhalten wir die Variablenreihenfolge A, B, C, D, wobei B und C vertauschbar sind, da sie nach Modell auf gleicher Kausalstufe stehen (vgl. Abb. 2a). Auf die Erläuterung der Pfadanalyse für nicht rekursive Kausalmodelle (vgl. Abb. 2b), möchte ich in dieser Arbeit verzichten.

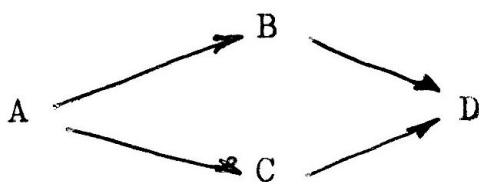


Abb. 2a

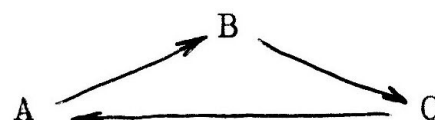


Abb. 2b

2) Wir berechnen die einfachen Korrelationskoeffizienten und bilden die Vierer-Matrix R, die durch die

Submatrix Q und den Spaltenvektor a untergliedert wird:

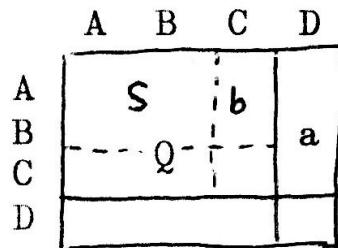


Abb. 3

3) Durch die folgende Matrix-Gleichung erhalten wir die drei ersten Pfadkoeffizienten (β_{AD} , β_{BD} , β_{CD}):

$$\beta = Q^{-1} \cdot a$$

Auf die ins Detail gehende Matrixberechnung (Matrixinversion und -multiplikation) möchte ich nicht weiter eingehen.

4) Derselbe Rechengang wird nun für C als abhängige und A und B als unabhängige Variablen durchgeführt. Wir betrachten dieses Mal Q als Ausgangsmatrix und unterteilen sie in die Submatrix S und den Spaltenvektor b (vgl. Abb. 3). Die Pfadkoeffizienten (β_{AC} , β_{BC}) erhalten wir wiederum durch die Matrixgleichung: $\beta = S^{-1} \cdot b$

5) Der letzte Pfadkoeffizient β_{AB} ist identisch mit r_{AB} .

Wir können nun die Pfadanalyse für eine oder mehrere konkurrierende Reihungen der Variablen wiederholen.

3. Die Voraussetzungen der Pfadanalyse

- Es muss eine schwache kausale Ordnung unter den Variablen bestehen.
- Die Beziehungen unter diesen Variablen sollen kausal geschlossen sein.
- Die Variablen sollen in unstandardisierten Werten vorliegen.
- Die Variablen werden auf dem Niveau der Intervall- oder Ratio-Skala gemessen.
- Die Variablen müssen normalverteilt sowie varianzhomogen sein.
- Es darf keine Multikollinearität bestehen.
- Die Variablen sollen linear verknüpft sein.
- Die Zuverlässigkeit der Messung der unabhängigen Variablen muss annähernd gleich gross sein.

Aus Zeitgründen ist es in dieser Arbeit nicht möglich auf alle aufgeführten Voraussetzungen einzugehen. Etwas ausführlicher sollen die für die Pfadanalyse besonders bedeutsamen Voraussetzungen a) und b) in Kapitel 4 behandelt werden. Auch auf das Problem - standardisierte versus unstandardisierte Variablen - wird in Kapitel 5 etwas detaillierter eingegangen.

Zu d):

Die Pfadanalyse lässt sich nicht nur mit Variablen, die auf der Intervall- oder Ratio-Skala gemessen

werden, durchführen. Sie kann auch mit unabhängigen oder sogar mit einer abhängigen nominalen Variablen ausgeführt werden. Allerdings ist dann eine besondere Transformation der nominalen Variablen notwendig, auf die ich hier aber nicht näher eingehen möchte.

Zu e):

Um schätzen zu können, d.h. um von einer Stichprobe auf eine Grundgesamtheit schliessen zu können, muss die Bedingung der Varianzhomogenität erfüllt sein. Verschiedene Autoren empfehlen, auf einen Varianzhomogenitätstest zu verzichten, da diese Bedingung bei empirischen Daten äusserst selten verletzt werde. Um Vertrauensintervalle für die Pfadkoeffizienten zu berechnen, d.h. um Signifikanzen angeben zu können, muss zusätzlich die Bedingung der Normalverteilung erfüllt sein. Das ist aber bei empirischen Daten häufig nicht der Fall; man muss dann entweder auf einen Signifikanztest verzichten, oder versuchen die Daten so zu transformieren, dass ein Normalverteilungstest positiv verläuft.

Zu f), g) und h) :

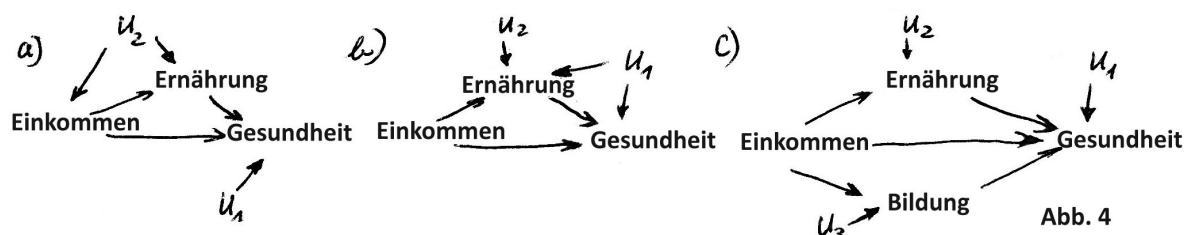
Wird die Inversion der Korrelationsmatrix Q der unabhängigen Variablen verunmöglicht, weil deren Determinante 0 wird (Multikollinearität), so kann die Pfadanalyse nicht durchgeführt werden. Trifft dieser unglückliche Fall ein, was extrem selten ist, so kann die für die Null zuständige Variable mit einem speziellen Verfahren identifiziert und geändert werden. In der Regel aber genügt es einige zufällig herausgesuchte Korrelationskoeffizienten an der 2. und 3. Kommastelle zu ändern. Auf die Bedingungen der linearen Verknüpfungen der Variablen und der Zuverlässigkeit der Messung bei unabhängigen Variablen möchte ich nicht näher eingehen.

4. Kausale Geschlossenheit und kausale Ordnung

Um herauszufinden, was die Pfadanalyse als Methode zur Erklärung soziologischer Phänomene leistet, ist es notwendig, die Voraussetzungen der kausalen Geschlossenheit und der kausalen Ordnung zu verstehen (Punkte a) und b)). Nur durch die genaue Vorabklärung dieser beiden Punkte, scheint mir eine sinnvolle Anwendung dieser Methode überhaupt erst möglich.

4.1 Kausale Geschlossenheit

Betrachten wir als erstes ein Beispiel aus der Soziologie: Man möchte zum Beispiel die Gesundheit der Bevölkerung, ausgedrückt durch das Mass Lebenserwartung, in den Entwicklungsländern erklären. Um etwas zu erklären, muss man nach den Gründen oder Ursachen forschen. Eine ‚Ursache‘ könnte die Höhe des Pro-Kopf-Einkommens der Bevölkerung sein. Man erkennt jedoch schnell, dass die Gesundheit noch durch spezifischere Mechanismen beeinflusst wird. So spielt hier sicher die Ernährung eine grosse Rolle. Die Ernährung ist jedoch keine unabhängige Variable, sondern hängt vom Einkommen ab. Sie hat in diesem Falle die Funktion einer intervenierenden Variablen (vgl. Abb. 4c). Dasselbe lässt sich von der neu eingeführten Variablen Bildung sagen.



Nun setzt aber die Methode der Pfadanalyse kausale Geschlossenheit der Beziehungen der Systemvariablen voraus. Das heisst in diesem Fall, dass zwischen U_1 , U_2 (übrigbleibende, unerklärte

Gründe) und den restlichen Systemvariablen nur bestimmte Kausalbeziehungen zulässig sind. So wäre z.B. in Abb. 4a oder 4b die Bedingung der kausalen Geschlossenheit nicht erfüllt. In Abb. 4c ist das System kausal geschlossen. Allgemein kann man sagen, dass die unerklärten Variablen oder auch Restvariablen U_1, U_2 , usw. voneinander unabhängig sein müssen. Dazu ein Beispiel: In unserem System scheint das Einkommen eine unabhängige Variable zu sein. Tatsächlich könnte sie jedoch von der Industrie oder Landwirtschaftsproduktion des betreffenden Landes abhängen. Angenommen diese Situation wäre gegeben, dann müssten wir jedoch diese neuen Variablen erst dann ins System einbeziehen, wenn wir einen unmittelbaren Zusammenhang zu den Variablen Ernährung oder Gesundheit vermuteten. Aus diesem Grund ist es z.B. sinnvoll, die Variable Bildung einzubeziehen, da diese nicht nur die Gesundheit beeinflusst, sondern auch vom Einkommen gesteuert wird (vgl. Abb. 4c) Durch die Anwendung der Pfadanalyse können wir die Stärke des direkten Einflusses der Restvariablen berechnen und erhalten so ein Maß dafür, ob unser Modell effektiv passt:

$$\beta_{U_1} = 1 - \sum_{i=1}^3 \beta_{iG} \cdot r_{iG}$$

In der Soziologie ist es jedoch häufig der Fall, dass die unerklärte Varianz sehr hoch ist. In unserem Beispiel wird die Gesundheit zu 90%, aber die Bildung nur zu 10% durch das Modell erklärt. Die 10% deuten an, dass unser System nicht ganz geschlossen ist.

Zusammenfassend kann man festhalten: Nur wenn man das Prinzip der kausalen Geschlossenheit richtig verstanden hat, kann man ein effektives und wirklichkeitsnahes Modell bauen. Ob ein System dann wirklich passt, lässt sich für das ganze System und für die einzelnen Variablen mit Hilfe der Pfadanalyse in der Restvariablenberechnung feststellen.

4.2 Kausale Ordnung

Beim Erforschen der Ursachen der übrigen Variablen stoßen wir auf ein interessantes Problem: Beeinflussen Gesundheit, Ernährung und Bildung nicht auch das Einkommen? Ist ein gesunder Mensch nicht auch eher in der Lage sich zu bilden?

Könnte man diese Fragen bejahen, so müsste man unser Modell abändern (vgl. Abb. 5). Unser System würde sich in ein nichtrekursives System, in ein sehr komplexes Netzwerk verwandeln. Um dieses Problem zu lösen, muss der Begriff Kausalität erklärt werden. Ein weiterer Grund für die Behandlung dieses Themas ergibt sich aus der Tatsache, dass seit Duncan 1966 seine soziologischen Beispiele

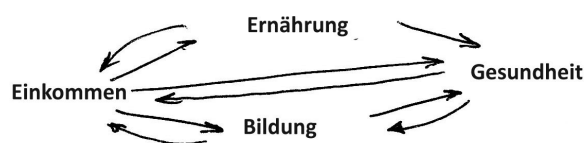


Abb.5

der Pfadanalyse veröffentlichte, die Konstruktion kausaler, rekursiver Modelle immer beliebter, um nicht zu sagen Mode geworden ist.

Es ist allgemein bekannt, dass die Ursache einer Folge vorangehen muss. Die Folgen ereignen sich nicht vor ihrer Ursache. Kausalität impliziert also den Zeitbegriff. So kann z.B. ein kranker, armer Mensch nur gesund werden, wenn er sich Nahrungsmittel und eine ärztliche Behandlung leisten kann. Wird dieser Mann mit Geld unterstützt, so stellt man nach einer gewissen Zeit fest, dass er gesund geworden ist. Der Kranke hat sich in einen Gesunden verwandelt, aufgrund des erhöhten Einkommens. Das Objekt unserer Untersuchung, der kranke Mann, hat sich also mit der Zeit verändert (gesunder Mann). Dieser gesunde Mann kann nun selbst sein Einkommen erhöhen, indem er z.B. eine Schule besucht,

Sprachen lernt und dadurch eine Stelle in der Stadt als Beamter annehmen kann.

Im allgemeinen befasst man sich in der Soziologie nicht mit einzelnen Individuen, sondern man fasst diese in Gruppen zusammen. Wenn wir nun alle Individuen (Kranke und Gesunde) eines Landes zu einer Gruppe zusammenfassen, so ist es einleuchtend, dass für die Gruppe das Einkommen auf die Gesundheit und die Gesundheit wiederum auf das Einkommen wirkt (Wechselwirkung).

Um eine eindeutige Kausalrichtung von Einkommen zu Gesundheit zu erhalten, muss man das Objekt der Untersuchung etwas spezifizieren. Es soll nämlich nur alle Individuen enthalten, die einen niedrigen Gesundheitsstand aufweisen. Wir haben dies in unserem Beispiel getan, indem wir uns nur auf die Entwicklungsländer beschränkten, in denen diese Annahme mehr oder weniger zutrifft.

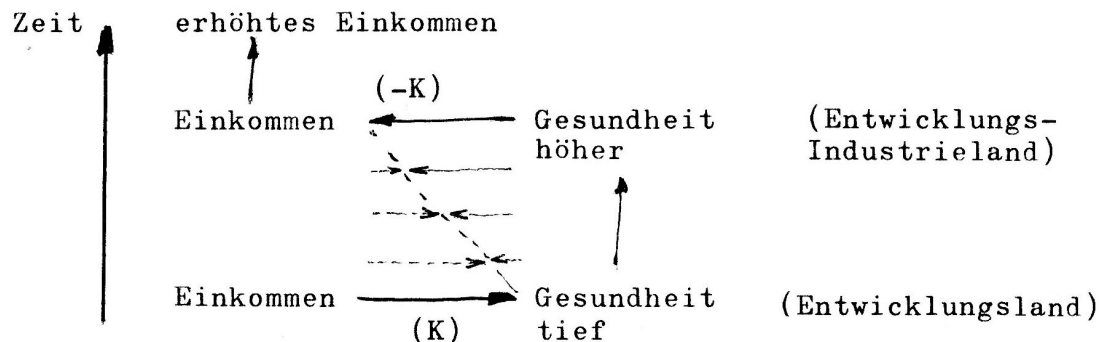


Abb.6

In Abb. 6 habe ich versucht den Entwicklungsprozess der Kausalrichtung (K ----> -K) und der Veränderung des Untersuchungsobjekts von tiefer Gesundheit zu erhöhter Gesundheit darzustellen.

Bei soziologischen Untersuchungen ist es häufig der Fall, dass man bei einer Querschnittanalyse z.B. das Einkommen und die Gesundheit zur gleichen Zeit betrachtet. Man konzentriert praktisch den Kausalprozess in einen Zeitpunkt. Dies darf man ohne weiteres tun, man soll sich jedoch nicht wundern, wenn man z.B. 1970 in einem Ölland hohe Einkommen und einen tiefen Gesundheitsstand feststellt (theoretisch würde man ja hohe Einkommen - hohe Gesundheit erwarten).

Welche Schlüsse lassen sich nach dieser Betrachtung von Kausalität für die Pfadanalyse ziehen?

- 1) Möchte man ein Mehrvariablensystem untersuchen und anschliessend eine Pfadanalyse für rekursive Systeme durchführen, so soll man als erstes das Kausalproblem gründlich studieren. Man soll sich der zeitlichen Veränderung der Variablen, sowie der zeitlichen Veränderung der Kausalrichtungen bewusst sein, und deshalb die Begriffe der Variablen so bestimmen, dass eindeutige Kausalrichtungen festgelegt werden können. Ist das zwischen einzelnen Variablen nicht möglich, so kann auch die komplexere Pfadanalyse für nicht-rekursive Systeme durchgeführt werden.
- 2) Erst wenn das Kausalitätsproblem gelöst ist, kann die Pfadanalyse durchgeführt werden. Dort wird jedoch lediglich noch die momentane Stärke der direkten Einflüsse auf die verschiedenen momentanen Variablen festgestellt. Eventuell ist es sogar möglich einige Pfade zu eliminieren, doch am Gesamtcharakter des Modells ändert sich dadurch nicht viel.
- 3) Die Effizienz der Pfadanalyse hängt deshalb im wesentlichen von der Qualität des Modells ab. Die Probleme der Messung, Standardisierung, usw. sollen zwar sorgfältig beachtet werden, doch sind sie bei der Anwendung dieser Methode Probleme zweiter Ordnung.

5. Standardisierung versus Nichtstandardisierung

In diesem Kapitel möchte ich etwas näher auf Voraussetzung c) (vgl. Kap. 3) eingehen.

Bei soziologischen Beispielen, in denen die Pfadanalyse durchgeführt wurde, stellt man häufig fest, dass standardisierte anstatt unstandardisierte Pfadkoeffizienten verwendet werden. Der Grund liegt in der Einfachheit der Eigenschaften des standardisierten β -Werts. Durch die Verwendung von standardisierten Koeffizienten lassen sich aber keine allgemeinen sozialen Gesetze entdecken. Der standardisierte Pfadkoeffizient geht durch folgende Gleichung aus dem nicht standardisierten Regressionskoeffizienten b hervor:

$$\beta = \frac{\sigma_x}{\sigma_y} \cdot b, \text{ wobei } \sigma_x: \text{Standardabweichung der unabhängigen Variablen}$$
$$\sigma_y: \text{Standardabweichung der abhängigen Variablen}$$

Da σ_x und σ_y je nach Population variieren, wird sich dementsprechend der standardisierte β -Wert je nach Population ändern. Um das oben beschriebene Vier-Variablenmodell zu verallgemeinern, müsste man es nicht nur an der Population der Entwicklungsländer, sondern auch an der Population der Industrieländer überprüfen. Es ist nun leicht einzusehen: Auch wenn das Kausalmodell für die Industrieländer zutreffen würde, so würde diese Verallgemeinerung oder dieses Gesetz durch die standardisierten β -Werte nicht aufgedeckt. Der Grund ist darin zu sehen, dass die Standardabweichungen der verschiedenen Populationen allzu unterschiedlich sind. So ist z.B. die Standardabweichung für das Pro-Kopf-Einkommen in den Industrieländern fast 100 Mal grösser als diejenige der Entwicklungsländer (1975)!

Es ist deshalb die Verwendung von unstandardisierten b -Werten zu empfehlen.

In den Naturwissenschaften wird zum Beispiel sehr selten mit standardisierten Daten gearbeitet. Nun könnte man natürlich einwenden, dass in der Soziologie, im Gegensatz zu den Naturwissenschaften, die Masse und Masseinheiten noch nicht allseits anerkannt sind. So gibt es zum Beispiel keinen allgemeinen Schichtindex; es fehlen auch allgemein normierte Masse um Einstellungen zu messen. Doch ich finde, dass dies ein Manko der Soziologie darstellt. Eine Wissenschaft kann sich nur dann entwickeln, und damit mehr allgemeine Gesetze entdecken, wenn sich die Soziologen auf bestimmte Masse einigen werden.

6. Klausurthema

Die Pfadanalyse als Mittel zur Erforschung kausaler Zusammenhänge in der Soziologie

- Kurze Einführung in die Pfadanalyse
- Abgrenzung der Pfadanalyse von anderen Methoden. Welche Voraussetzungen sind nötig, um diese Methode anzuwenden?
- Erklärung der Pfadanalyse anhand eines oder mehrerer Beispiele aus den Sozialwissenschaften (eventuell ein Vergleich mit einem Beispiel aus den Naturwissenschaften).
- Interpretation der Ergebnisse. Was kann die Pfadanalyse erklären und was nicht?
- Gibt es andere oder bessere Verfahren um Annahmen über Kausalzusammenhänge nachzuweisen?

Literaturverzeichnis:

- Blalock, H.M.jr. (Hrsg.): Causal Models in the Social Sciences.
Atherton 1971
- Boudon, R./Lazarsfeld, P. L'analyse empirique de la causalité. (Hrsg.):
Paris 1966
- Galtung, J.: Theorie and Methods of Social Research.
Oslo 1967
- Holm, K.(Hrsg.): Die Befragung 5: Pfadanalyse, Coleman-Verfahren.
München 1977
- Hummel, H.J./Ziegler, R,. Korrelation und Kausalität, Band 1 - 3. (Hrsg.):
Stuttgart 1976
- Marinell, G.: Multivariate Verfahren.
München Wien 1977
- Nie, N.H.e.a.: Statistical Package for the Social Sciences.
USA 1975
- Van de Geer, J.P.: Introduction to Multivariate Analysis for the Social Sciences.
San Francisco 1971