

**Pfadanalyse
Ein statistisches Verfahren
zur Untersuchung linearer Kausalmodelle**

H. DIETER SEIBEL

G.T. NYGREEN

Department of Sociology, Princeton University

Die Pfadanalyse ist ein statistisches Verfahren zur Überprüfung linearer Kausalmodelle. Die Stärke der kausalen Beziehung zwischen je zwei Variablen wird durch Pfadkoeffizienten ausgedrückt. Pfadkoeffizienten lassen sich aus Korrelationskoeffizienten ableiten oder als partielle Regressionskoeffizienten aus standardisierten Eingabedaten berechnen. Die erste, hier ausführlich vorggeführte Methode lässt sich von Hand durchführen. Die Stärke des Einflusses sämtlicher nicht spezifizierten Variablen wird durch Residualpfadkoeffizienten wiedergegeben; ihr Quadrat ergibt den Prozentsatz der nicht erklärten Varianz.

Path analysis is an approach for evaluating causal models which can be specified as recursive systems of linear equations. The strength of the association between any causally related pair of variables is expressed by a path coefficient which statistically is identical to a standardized partial regression coefficient. The system of path coefficients can be computed by hand from the simple correlation matrix (zero-order correlation coefficients). The strength of the influence of all nonspecified variables is given by a residual path coefficient. The square of this residual represents the proportion of variance unexplained in the causal model.

EINFÜHRUNG

Die Pfadanalyse (*path analysis*) ist ein statistisches Verfahren zur Erforschung komplexer Zusammenhänge. Sie wird primär zur Überprüfung linearer Kausalmodelle angewandt. Sie wurde 1934 von SEWALL WRIGHT in seinem Aufsatz «The Method of Path Coefficients» in die Statistik eingeführt. Wegen des enormen Aufwandes an Rechenarbeit, der zur Kalkulation der Pfadkoeffizienten erforderlich ist, blieb die Methode jahrzehntelang unangewandt. Erst mit der Verbreitung elektronischer Großrechenanlagen tauchte sie wieder aus der Vergessenheit auf. In die Soziologie eingeführt wurde sie in den letzten Jahren durch die Arbeiten von

BLALOCK (1964), BLAU & DUNCAN (1967), BOYLE (1970), BOUDON (1965), DUNCAN (1966), GOLDBERGER (1970), HEISE (1969) und LAND (1969). Seitdem hat sich ihre Anwendung stetig

verbreitert. In jüngster Zeit wurde sie an einigen amerikanischen Universitäten in Computer-Paketprogramme wie SPSS und DATA-TEXT aufgenommen. Da die Methode in der europäischen Soziologie noch wenig bekannt ist und auch in der amerikanischen Literatur nirgends *umfassend* dargestellt ist, wird im folgenden eine systematische Einführung in die Berechnung von Pfadkoeffizienten gegeben. *Diese Einführung soll den Leser in die Lage versetzen, Pfadkoeffizienten selbst zu berechnen, wenn die Matrix der Korrelationskoeffizienten gegeben ist.*

Bei der Pfadanalyse wird ein mehr oder weniger komplexes Netz von Beziehungen zwischen unabhängigen und abhängigen Variablen analysiert, wobei die Beziehungen in der graphischen Darstellung durch Pfeile angegeben sind. Jeder Pfeil steht für einen sogenannten Pfad (*path*), dem ein Pfadkoeffizient zugeordnet ist. Pfade in der Grundgesamtheit werden mit Beta

bezeichnet, Pfade in der Stichprobe mit b oder p . Der Betawert gibt die Stärke des Zusammenhangs zwischen den jeweiligen Variablen unter Berücksichtigung des Einflusses aller übrigen in die Berechnung eingeführten Variablen an. Der Pfadkoeffizient oder Betawert ist ein *standardisierter partieller Regressionskoeffizient*. Da Pfadkoeffizienten standardisiert sind, sind sie direkt miteinander vergleichbar. – Zu jeder abhängigen Variablen führt außerdem der Residualpfad p_{res} : der Pfadkoeffizient aller nicht erklärten Variablen. Beziehungen zwischen unabhängigen Variablen werden durch den Korrelationskoeffizienten r dargestellt (in den Abbildungen 1, 2 und 3 als Kurve mit zwei Pfeilspitzen wiedergegeben).

Voraussetzung für die Anwendbarkeit der Pfadanalyse ist kardinale, mindestens aber ordinale Messbarkeit der Daten, außer bei dichotomen Variablen¹

Anleitung zur Berechnung von Pfadkoeffizienten

VERFAHREN 1:

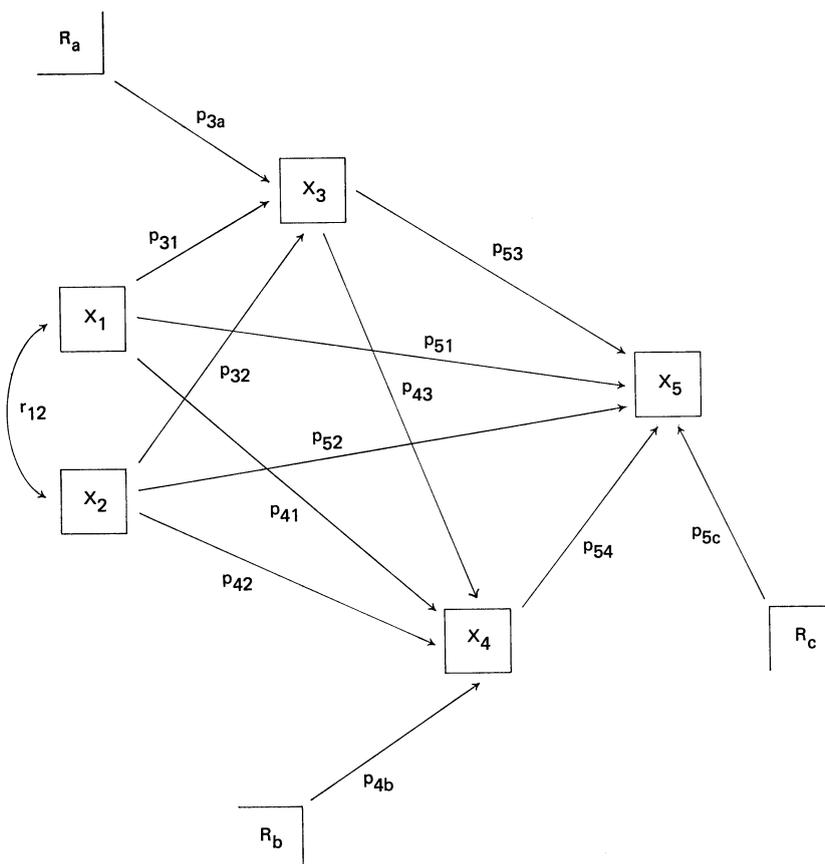


Abbildung 1: Pfaddiagramm mit sämtlichen Pfaden

¹ Berechnet man die Pfadkoeffizienten aus den Rohdaten unter Verwendung der Regressionsanalyse, so lassen sich auch nicht-dichotome, nominal gemessene Daten verwenden. Dabei wird jede Komponente der nominal messbaren Variablen in eine sogenannte Scheinvariable (dummy variable) umgewandelt, die aus den zwei Komponenten „Merkmal trifft zu“ und „Merkmal trifft nicht zu“ besteht. Im weiteren werden die Scheinvariablen wie alle anderen Variablen behandelt. - Zur Frage der Verwendbarkeit ordinaler Daten siehe LABOVITZ (1970)

Gehen wir von einem Kausalmodell aus, wie es in Abbildung 1 dargestellt ist. x_1 und x_2 sind zwei unabhängige Variablen. x_5 ist eine abhängige Variable, und x_3 und x_4 sind abhängig in bezug auf x_1 und x_2 und unabhängig in bezug auf x_5 . Das Modell ist vollkommen rekursiv, d.h. die Pfade sämtlicher logisch konsistenten uni-linearen Kausalbeziehungen sind angegeben. Die Gesamtheit der Gleichungen eines Kausalmodells wird als *Strukturmodell* bezeichnet. Drei der Beziehungen innerhalb des Modells (Abb. 1) können in Form von multiplen Regressionsgleichungen ausgedrückt werden:

$$x_3 = b_{32}x_2 + b_{31}x_1 + e_3 \quad (1)$$

$$x_4 = b_{43}x_3 + b_{42}x_2 + b_{41}x_1 + e_4 \quad (2)$$

$$x_5 = b_{54}x_4 + b_{53}x_3 + b_{52}x_2 + b_{51}x_1 + e_5 \quad (3)$$

Da Pfadkoeffizienten standardisierte partielle Regressionskoeffizienten sind, sind die Regressionskoeffizienten für die strukturellen Gleichungen 1–3 gleichzeitig Pfadkoeffizienten, wenn die Variablen in Standardform gemessen werden. Pfadkoeffizienten werden häufig aus Rohdaten berechnet, wobei Regressionsprogramme benutzt werden, die partielle Koeffizienten aus standardisierten Eingabedaten berechnen. Die Berechnung ist ohne Computer kaum durchführbar.

VERFAHREN 2:

Die strukturellen Gleichungen lassen sich algebraisch so umformen, dass jeder Pfadkoeffizient als Ausdruck von Korrelationskoeffizienten und anderen Pfadkoeffizienten wiedergegeben wird. DUNCAN (1966, S. 5) bezeichnet die allgemeine Form solcher Ausdrücke als *Grundtheorem der Pfadanalyse*. Wenn die Korrelationskoeffizienten bekannt sind, lassen sich mit Hilfe des *Grundtheorems* die Pfadkoeffizienten berechnen. Bei dieser Methode wird das Regressionsverfahren also umgangen. Dieses Verfahren wird hier vorgeführt, weil es die Von-Hand-Berechnung ermöglicht.

Das *Grundtheorem der Pfadanalyse* lautet:

$$r_{ij} = \sum_q p_{iq} r_{qj} \quad (4)$$

wobei der Index q durch alle Variablen läuft, von denen Pfade direkt zu x_j führen. Dieses *Grundtheorem* stellt in allgemeiner Form die Gesamtheit der nichthomogenen Gleichungen dar, von denen jede für die Berechnung eines Pfadkoeffizienten gelöst werden kann. Die Korrelation zwischen je einer abhängigen Variablen (z. B. x_5) und den zugehörigen unabhängigen Variablen (x_1, x_2, x_3, x_4) lässt sich nun in Form des *Grundtheorems* (4) als Funktion der Pfadkoeffizienten und weiterer Korrelationskoeffizienten darstellen:

$$r_{51} = p_{51}r_{11} + p_{52}r_{21} + p_{53}r_{31} + p_{54}r_{41} \quad (5)$$

$$r_{52} = p_{51}r_{12} + p_{52}r_{22} + p_{53}r_{32} + p_{54}r_{42} \quad (6)$$

$$r_{53} = p_{51}r_{13} + p_{52}r_{23} + p_{53}r_{33} + p_{54}r_{43} \quad (7)$$

$$r_{54} = p_{51}r_{14} + p_{52}r_{24} + p_{53}r_{34} + p_{54}r_{44} \quad (8)$$

In Matrixform ergeben diese Gleichungen:

$$\begin{bmatrix} r_{11} & r_{12} & r_{13} & r_{14} \\ r_{21} & r_{22} & r_{23} & r_{24} \\ r_{31} & r_{32} & r_{33} & r_{34} \\ r_{41} & r_{42} & r_{43} & r_{44} \end{bmatrix} \begin{bmatrix} p_{51} \\ p_{52} \\ p_{53} \\ p_{54} \end{bmatrix} = \begin{bmatrix} r_{51} \\ r_{52} \\ r_{53} \\ r_{54} \end{bmatrix}$$

A x b

Somit ist das Problem der Bestimmung von Pfadkoeffizienten für die vier Pfade nach x_5 in diesem Modell formalisiert. Die Lösung für den Vektor x ergibt sich aus unbekanntem Pfadkoeffizienten: Da die Matrix A aus empirischen Daten abgeleitete Korrelationskoeffizienten enthält und die Werte des Vektors b aus den gleichen Daten berechnet werden, bleiben die Pfadkoeffizienten die einzigen Unbekannten.

Wenn A die Matrix der bekannten Korrelationskoeffizienten ist, x ein Vektor von bestimmten Pfadkoeffizienten, die sich aus dem *Grundtheorem* ergeben, und b ein weiterer Vektor, der sich aus bekannten Korrelationskoeffizienten ableiten lässt, dann lässt sich die Gesamtheit aller simultanen Gleichungen in der allgemeinen Matrixform

$$Ax = b \tag{9}$$

darstellen. Durch Umformung ergibt sich

$$x = A^{-1}b \tag{10}$$

Der Vektor x ist also gleich dem Produkt aus dem Vektor b und dem Kehrwert der Matrix A . In ähnlicher Weise lassen sich die Pfadkoeffizienten ableiten, die zu jeder der drei abhängigen Variablen führen (Abb. 1).

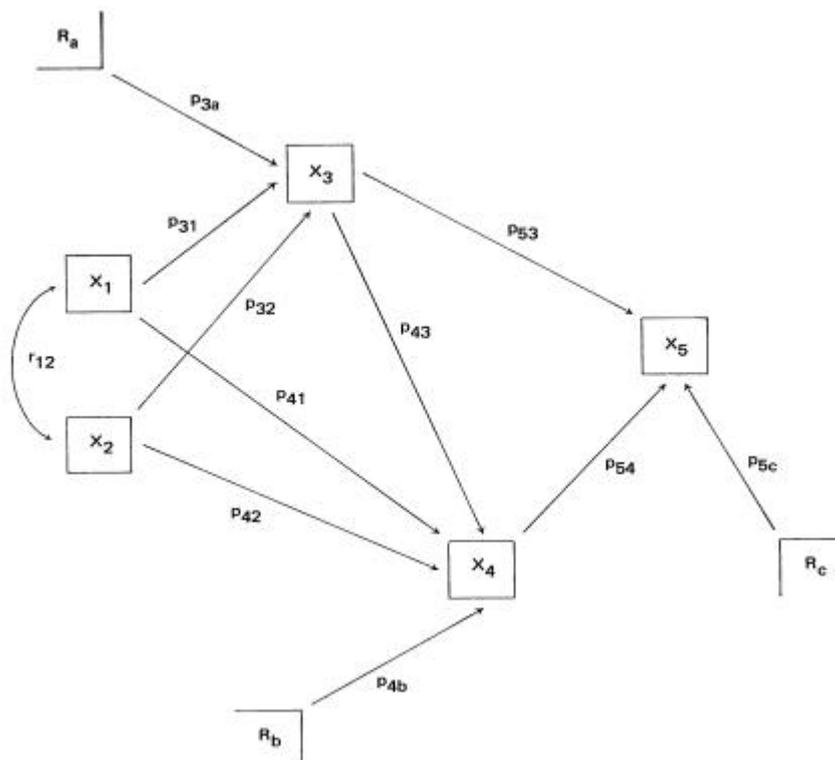


Abbildung 2: Pfaddiagramm mit zwei direkten Pfaden zu x_5

Im allgemeinen enthalten die Kausalmodelle, mit denen man im konkreten Fall arbeitet, nur einen Teil der möglichen Pfade (Abb. 2), so dass das obige Berechnungsschema nicht anwendbar ist. In Abbildung 2 sind zwei der Pfade zu x_5 nicht angegeben². Damit werden die Pfadkoeffizienten für die eliminierten Pfade in den Gleichungen 5-8 Null, so dass jeder P_{51} - und P_{52} - Ausdruck entfällt:

$$r_{51} = p_{53}r_{13} + p_{54}r_{14} \quad (11)$$

$$r_{52} = p_{53}r_{23} + p_{54}r_{24} \quad (12)$$

$$r_{53} = p_{53}r_{33} + p_{54}r_{34} \quad (13)$$

$$r_{54} = p_{53}r_{34} + p_{54}r_{44} \quad (14)$$

Da nur zwei Unbekannte zu bestimmen sind, werden nur zwei Gleichungen benötigt. Würden alle vier Gleichungen verwandt, so ergäben sich sechs Zweier-Kombinationen und somit sechs Werte für jeden Koeffizienten. Da die Pfade p_{54} und p_{53} bestimmt werden sollen, werden nur die Gleichungen (13) und (14) verwandt. Die Lösung ergibt sich aus (10) und (4) wie oben dargestellt.

ZAHLENBEISPIEL

Im folgenden wird nun ein Zahlenbeispiel für die Berechnung der Pfadkoeffizienten p_{54} und p_{53} vorgeführt, das von hypothetischen Korrelationen zwischen den in Abbildung 2 angegebenen Variablen ausgeht. Die hypothetische Korrelationsmatrix für die fünf Variablen soll lauten:

1,00	0,12	0,41	0,42	0,29
0,12	1,00	0,35	0,15	0,31
0,41	0,35	1,00	0,56	0,52
0,42	0,15	0,56	1,00	0,49
0,29	0,31	0,52	0,49	1,00

Für die Berechnung der beiden Koeffizienten der Pfade zu x_5 sind nun die Gleichungen nach (13) und (14) zu erstellen. In der Form der Gleichung (9) ausgedrückt ergibt sich:

$$\begin{matrix} \begin{bmatrix} 1,00 & 0,56 \\ 0,56 & 1,00 \end{bmatrix} & \begin{bmatrix} p_{53} \\ p_{54} \end{bmatrix} & = & \begin{bmatrix} 0,52 \\ 0,49 \end{bmatrix} \\ A & x & & b \end{matrix}$$

Der Kehrwert von A ist

$$A^{-1} = \begin{bmatrix} 1,457 & -0,816 \\ -0,816 & 1,457 \end{bmatrix}$$

² Zwei alternative Gründe lassen sich anführen: Entweder sind diese Pfade nach dem theoretischen Grundmodell nicht sinnvoll; oder sie sind sinnvoll, haben sich bei der Berechnung aber als nichtsignifikant erwiesen.

Durch Einsetzen in (10) ergibt sich

$$A^{-1}b = \begin{bmatrix} 0,358 \\ 0,290 \end{bmatrix}$$

Damit sind die Pfadkoeffizienten

$$P_{53} = 0,358 \text{ und}$$

$$P_{54} = 0,290$$

Zur Berechnung der Koeffizienten der Pfade zu x_4 (Abb. 2) gehen wir von unserem *Grundtheorem* aus:

$$r_{41} = p_{41}r_{11} + p_{42}r_{21} + p_{43}r_{31} \quad (15)$$

$$r_{42} = p_{41}r_{12} + p_{42}r_{22} + p_{43}r_{32} \quad (16)$$

$$r_{43} = p_{41}r_{13} + p_{42}r_{23} + p_{43}r_{33} \quad (17)$$

oder in Matrixform:

$$\begin{bmatrix} r_{11} & r_{21} & r_{31} \\ r_{12} & r_{22} & r_{32} \\ r_{13} & r_{23} & r_{33} \end{bmatrix} \begin{bmatrix} p_{41} \\ p_{42} \\ p_{43} \end{bmatrix} = \begin{bmatrix} r_{41} \\ r_{42} \\ r_{43} \end{bmatrix}$$

$$A \quad x \quad b$$

Durch Einsetzen der Zahlenwerte für die Korrelationskoeffizienten ergeben sich die drei Matrices

$$\begin{bmatrix} 1,00 & 0,12 & 0,41 \\ 0,12 & 1,00 & 0,35 \\ 0,41 & 0,35 & 1,00 \end{bmatrix} \begin{bmatrix} p_{41} \\ p_{42} \\ p_{43} \end{bmatrix} = \begin{bmatrix} 0,42 \\ 0,15 \\ 0,56 \end{bmatrix}$$

$$A \quad x \quad b$$

Der Kehrwert von A ist

$$A^{-1} = \begin{bmatrix} 1,203 & 0,032 & -0,504 \\ 0,032 & 1,140 & -0,412 \\ -0,504 & -0,412 & 1,351 \end{bmatrix}$$

Die Multiplikation mit dem Vektor b ergibt

$$A^{-1}b = \begin{bmatrix} 0,228 \\ -0,046 \\ 0,483 \end{bmatrix}$$

Damit sind die Pfadkoeffizienten

$$p_{41} = 0,228$$

$$p_{42} = -0,046$$

$$p_{43} = 0,483^3$$

DIE RESIDUEN

Bei empirischem Vorgehen wird es in keiner Wissenschaft je vorkommen, dass eine abhängige Variable völlig durch die spezifizierten unabhängigen Variablen bestimmt ist. Es bleibt ein Residuum von Variablen (im Extremfalle dem Messfehler gleich), die die Varianz der abhängigen Variablen bestimmen. Dieses Residuum ist in dem Modell multipler Regressionsgleichungen eingeschlossen, ebenso wie in dem Modell der Pfadkoeffizienten. Die Pfadkoeffizienten sind als b -Werte, die Residuen als e -Werte («error») wiedergegeben. Das strukturelle Modell in Abbildung 1 kann nun in der Form

$$x_3 = p_{32}x_2 + p_{31}x_1 + p_{3a}R_a \quad (18)$$

$$x_4 = p_{43}x_3 + p_{42}x_2 + p_{41}x_1 + p_{4b}R_b \quad (19)$$

$$x_5 = p_{54}x_4 + p_{53}x_3 + p_{52}x_2 + p_{51}x_1 + p_{5c}R_c \quad (20)$$

dargestellt werden. Das Residuum erklärt die Varianz der abhängigen Variablen, die nicht durch die spezifizierten unabhängigen Variablen erklärt wird. Daher handelt es sich bei den Gleichungen 18–20 um ein *geschlossenes* Strukturmodell: alle Faktoren, die die Varianz erklären, sind darin enthalten. Im folgenden wird nun gezeigt, dass der Residualpfad (18) tatsächlich die Gesamtheit aller nicht spezifizierten Pfade zu x_3 darstellt.

Die Korrelation einer abhängigen Variablen (i) mit sich selbst (i) kann in der Form

$$r_{ii} = 1,0 = \sum_q p_{iq}r_{iq} \quad (21)$$

dargestellt werden. Da der Residualpfad außerdem der Differenz zwischen der Gesamtheit sämtlicher Einflüsse (1,0) und den Beiträgen aller spezifizierten Pfade (q) gleich ist, gilt

$$P_{ia}^2 = 1 - \sum_q p_{iq}r_{iq} \quad (22)$$

³ Wenn die Berechnung mit einem Computer durchgeführt wird, ergeben sich leichte Abweichungen, da die Rundungsfehler geringer sind.

wobei der Index q durch alle Variablen läuft, von denen Pfade zu x_1 führen.

Aus dem ersten Zahlenbeispiel setzen wir nun die Korrelationskoeffizienten zwischen x_5 und den beiden unabhängigen Variablen

$$r_{53} = 0,52$$

$$r_{54} = 0,49$$

und die beiden Pfadkoeffizienten

$$p_{53} = 0,358$$

$$p_{54} = 0,290$$

in den Ausdruck (22) ein, wobei der Index i dem Index 5 und der Index q den Indices 3 und 4 entspricht:

$$p_{5c}^2 = 1 - (0,358)(0,52) - (0,290)(0,49)$$

$$= 0,672 (= \text{Anteil der nicht erklärten Varianz})$$

Der Residualpfad ist gleich der Quadratwurzel:

$$p_{5c} = 0,82^4$$

Als ein Beispiel aus der sozialwissenschaftlichen Forschung wird in Abbildung 3 das von BLAU & DUNCAN (1967) erstellte Pfaddiagramm dargestellt, das die Berufsposition der 20- bis 64 jährigen Amerikaner im März 1962 aus den Faktoren *Ausbildung des Vaters*⁵, *Beruf des Vaters*, *Ausbildung des Respondenten* und *Erste Berufsposition des Respondenten* erklärt. Die Koeffizienten wurden aus den Daten einer repräsentativen Stichprobe von 20700 Respondenten berechnet.

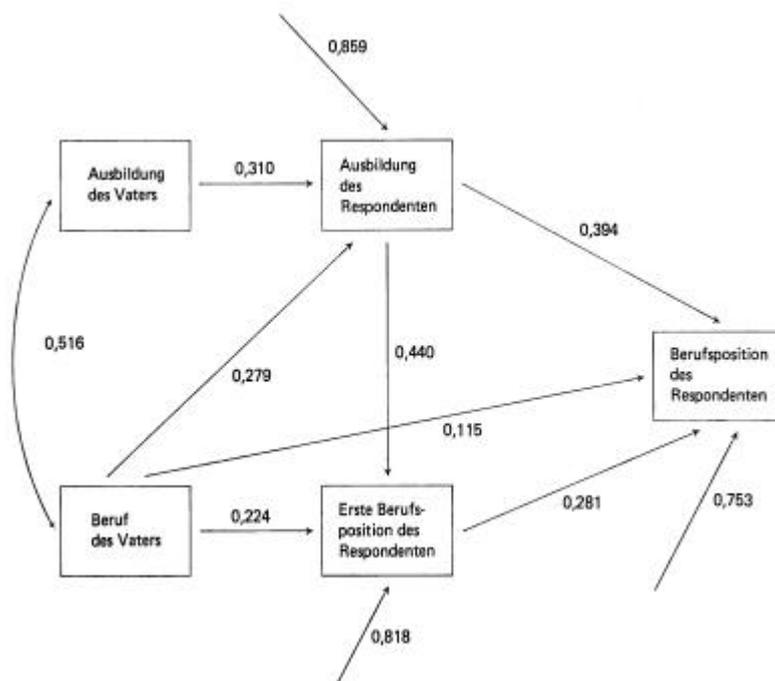


Abbildung 3: Blau und Duncans Pfaddiagramm

⁴ Es ist zu betonen, dass die Berechnung der Pfadkoeffizienten aus der Korrelationsmatrix nur *eine* Berechnungsmethode aus mehreren darstellt.

⁵ Der Pfad zwischen *Ausbildung des Vaters* und *Berufsposition des Respondenten* ist nicht signifikant und daher nicht eingezeichnet.

Man beachte die Übersichtlichkeit der Darstellung im Vergleich zu der sonst üblichen verwirrenden Vielzahl von zwei- und mehrdimensionalen Tabellen.

DER ANTEIL DER ERKLÄRTEN VARIANZ

Die folgende Darstellung geht nun nicht mehr von der Matrix der Korrelationskoeffizienten aus, sondern von der Regressionsanalyse. Die Berechnung von Hand lässt sich hier kaum noch durchführen. Es ist ratsam, einen Computer für die Berechnung heranzuziehen.

Der Pfadkoeffizient gibt die Steigung der Regressionsgeraden in standardisierter Form wieder. Ist die Streuung der Werte um die Regressionsgerade gering, so ist die durch den Pfadkoeffizienten ausgedrückte Erklärungskraft der unabhängigen Variablen stärker, als wenn die Werte weit gestreut sind. Der Pfadkoeffizient selbst sagt nichts über die Streuung der Werte um die Regressionsgerade aus. Es bedarf daher einer zweiten Messzahl, die sich bei der Berechnung der Pfadkoeffizienten aus der Regressionsanalyse als Nebenprodukt ergibt: des Prozentsatzes der erklärten Varianz (*percent variance explained*).

Der Prozentsatz der *insgesamt* erklärten Varianz ist $100 (1 - p_{res}^2)$ Setzt man den Wert des

Residualpfades in Abbildung 3 in diese Formel ein, so ergibt sich, dass Blau & Duncan $100 (1 - 0,753^2) = 100 (1 - 0,567) = 43,3$ Prozent der Varianz in der Berufsposition der Befragten durch die spezifizierten Variablen erklären. Nur bei dem Residualpfad besteht ein direkter Zusammenhang zwischen dem Pfadkoeffizienten und der (*insgesamt!*) erklärten Varianz.

Die unabhängigen Variablen, die zur Erklärung der abhängigen Variablen herangezogen werden, können in verschiedener Reihenfolge in die Berechnung eingeführt werden. Diese Reihenfolge ist für die Pfadkoeffizienten ohne Bedeutung. Dagegen spielt sie eine entscheidende Rolle bei der Berechnung der einzelnen Prozentsätze der Varianz in der abhängigen Variablen, die die einzelnen unabhängigen Variablen erklären. Hinsichtlich der Kriterien, nach denen die Reihenfolge der unabhängigen Variablen bestimmt wird, gibt es zwei Methoden: eine schritt-

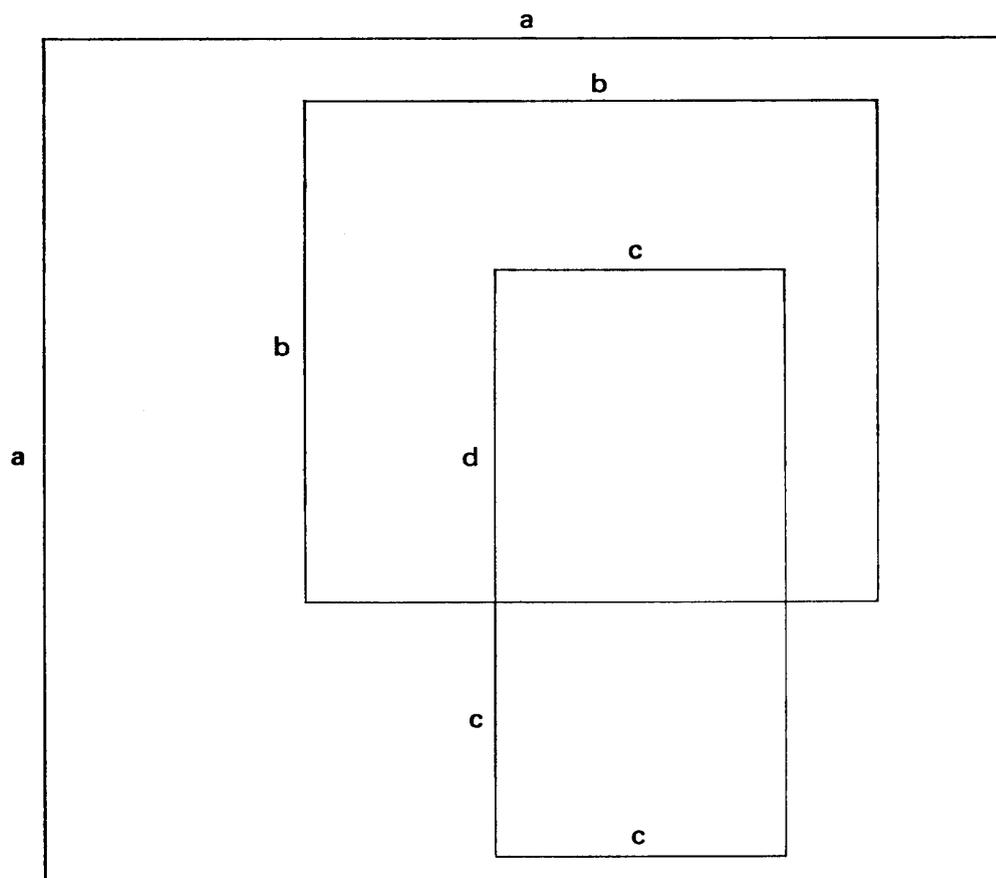
weise (*stepwise*) und eine multiple (*multiple*) Methode. Bei der schrittweisen Methode richtet sich die Reihenfolge nach der Signifikanz des Beitrages, den eine unabhängige Variable zur Erklärung einer abhängigen Variablen leistet. Die Variable wird als erste eingeführt, die den signifikantesten Beitrag leistet, und die übrigen Variablen folgen in der Reihenfolge abnehmender Signifikanz. Erweist sich der Beitrag einer Variablen als insignifikant (nach dem zugrunde-gelegten Signifikanzniveau), so wird der betreffende Pfad eliminiert; das gleiche gilt für die nachfolgenden Variablen, deren Signifikanz noch geringer ist. Bei exaktem Vorgehen sind meist zwei Rechengänge erforderlich: Im ersten wird festgestellt, welche unabhängigen Variablen insignifikant in ihrer Beziehung zur abhängigen Variablen sind. Diese werden für die zweite Berechnung eliminiert. In der zweiten Berechnung werden dann nur die Variablen eingeführt, die sich als signifikant erwiesen haben. Bei dieser Neuberechnung ergeben sich Abweichungen in den Werten der Pfadkoeffizienten, da nun der Einfluss der insignifikanten Variablen ausgeschaltet ist, der bei der ersten Berechnung noch von - wenn auch geringer - Bedeutung war.

Bei der multiplen Methode wird die Reihenfolge vorher festgelegt. Dabei gibt es wiederum zwei Möglichkeiten: Man kann eine Variable nach der anderen in die Berechnung einführen; oder man führt sie gleichzeitig ein. Beide Verfahren lassen sich miteinander kombinieren. Häufig werden die ersten zwei oder drei unabhängigen Variablen, die den Wissenschaftler auf Grund seines theoretischen Ausgangsmodells am meisten interessieren, nacheinander eingeführt, und die übrigen Variablen werden dann gleichzeitig eingeführt.

Welche der beiden Methoden vorgezogen wird, hängt von dem theoretischen Ausgangsmodell ab. In den meisten Fällen sind die Vorstellungen über die kausale Reihenfolge der unabhängigen Variablen so unvollständig⁶, dass die schrittweise Methode geraten erscheint. Da e Oft soll diese Reihenfolge durch die Pfadanalyse überhaupt erst bestimmt werden.

sich aber der Bereich der von mehreren Variablen erklärten Varianz im allgemeinen überschneidet (Abb. 4), empfiehlt sich oft die Anwendung der multiplen Methode in einem zweiten Rechengang. In diesem wird dann die Reihenfolge, die sich bei der schrittweisen Methode ergeben hat, für die Variablen umgekehrt, für die man den Überschneidungsbereich bestimmen will.

Die von zwei Variablen erklärte Varianz ist in Abbildung 4 graphisch dargestellt. a^2 gibt die insgesamt zu erklärende Varianz der abhängigen Variablen x_5 (Abb. 2) wieder. $b^2 + c^2$ geben die insgesamt von den beiden unabhängigen Variablen x_3 und x_4 erklärte Varianz wieder. b^2 stellt die durch x_3 im *schrittweisen Verfahren* erklärte Varianz dar, wobei x_3 als erste Variable in die Rechnung eingeführt wurde. c^2 stellt die durch x_4 im *schrittweisen Verfahren* erklärte Varianz dar, wobei x_4 als zweite Variable in die Rechnung eingeführt wurde. $c(c + d)$ stellt die durch x_4 im *multiplen Verfahren* erklärte Varianz dar, wobei x_4 als erste Variable in die Rechnung eingeführt wurde. $b^2 - cd$ stellt die durch x_3 im *multiplen Verfahren* erklärte Varianz dar, wobei x_3 als zweite Variable in die Rechnung eingeführt wurde. $a^2 - (b^2 + c^2)$ gibt die Residualvarianz wieder. Die Prozentsätze sind fiktiv.



⁶ Oft soll die Reihenfolge durch die Pfadanalyse überhaupt erst bestimmt werden.

$$1 \text{ cm}^2 = 1\%$$

$$a^2 = 100\%$$

$$b^2 + c^2 = 45\%$$

$$b^2 = 36\%$$

$$c^2 = 9\%$$

$$c(c + d) = 21\%$$

$$b^2 - c d = 24\%$$

$$c d = 12\%$$

Abbildung 4
Gesamtvarianz einer abhängigen Variablen und durch
zwei unabhängige Variablen erklärte Teilvarianzen

Literatur

- BLALOCK, H. M. Jr. 1964. Causal inferences in nonexperimental research. Chapel Hill: University of North Carolina Press.
- BLAU, P. M. & DUNCAN, O. D. 1967. The American occupational structure. New York: Wiley.
- BOYLE, R. P. 1970. Path analysis and ordinal data. *American Journal of Sociology* 75., 461-480.
- BOUDON, R. 1965. A method of linear causal analysis: Dependence analysis. *American Sociological Review* 30, 365-374.
- DUNCAN, O. D. 1966. Path analysis: Sociological examples. *American Journal of Sociology* 72, 1-16.
- GOLDBERGER, A. S. 1970. On Boudon's method of linear causal analysis. *American Sociological Review* 35, 97-101.
- HEISE, D. R. 1969. Problems in path analysis and causal inference. In: BOROATTA, E. (Hg.). *Sociological methodology*. San Francisco: Jossey-Bass. S.38-73.
- LABOWITZ, S. 1970. The assignment of numbers to rank order categories. *American Sociological Review* 35, 515-524.
- LAND, K. C. 1969. Principles of path analysis. In: BORGATTA, E. (Hg.). *Sociological methodology*. San Francisco: Jossey-Bass. S.3-37.
- WRIGHT, S. 1934. The method of path coefficients. *Annals of Mathematical Statistics* S. 161-215.