

1. Kapitel

Übersicht

Jürgen Bredenkamp

1. Zur Prüfung sog. Kausalhypothesen

Selten begnügt man sich in der Psychologie, wie in anderen empirischen Wissenschaftszweigen auch, damit, die korrelativen Beziehungen zwischen verschiedenen Variablen zu konstatieren, ohne zu prüfen, welche Variablen auf andere einwirken. Man intendiert Aussagen wie z.B. „X wirkt sich auf Y aus“. Derartige Interpretationen sind bekanntlich nicht einfach aus den korrelativen Beziehungen zu erschließen. Entweder stellt man aufgrund experimenteller Planungen sicher, daß die Variation von Y als Folge der Variation von X interpretiert werden kann, oder es wird, falls experimentelles Handeln nicht möglich ist, ein Modell über die Beeinflussungsrichtung zwischen den Variablen aufgestellt, das auch dann prüfbar ist, wenn es sich um reine Korrelationsforschung handelt.

Dem Experiment, dessen Planung und Auswertung ausführlich durch Hager und Westermann in diesem Band behandelt werden, kommt unter den verschiedenen Methoden zur Überprüfung sog. Kausalhypothesen - dieser Begriff wird noch präzisiert werden - eine Sonderstellung zu. Der Experimentator selbst stellt verschiedene Bedingungen her (er „manipuliert“ eine Variable) und beobachtet die Auswirkungen dieser „unabhängigen“ auf eine andere „abhängige“ Variable. Dadurch wird eine zeitliche Abfolge unabhängige Variable „X“ → abhängige Variable „Y“ hergestellt, und die korrelative Beziehung zwischen beiden Größen kann nicht derart interpretiert werden, daß sich Y auf X auswirkt. Allerdings reicht das bisher geschilderte Vorgehen noch nicht aus, um behaupten zu können, daß X auf Y einwirkt. Mit X könnten eine oder mehrere Störvariablen „St“ korreliert sein, die „in Wirklichkeit“ für den beobachteten Zusammenhang zwischen X und Y verantwortlich sind. Um die Möglichkeit einer derartigen Scheinbeziehung zwischen X und Y, wie sie in Abb. 1 dargestellt ist, zu reduzieren, bedient man sich in der experimentellen Psychologie verschiedener Kontrollverfahren; in der Terminologie Campbells und Stanleys (1963) kommt einem Experiment um so größere interne

Validität zu, je geringer die Möglichkeit zu derartigen Scheinbeziehungen ist. Wichtig ist vor allem die zufällige Zuweisung der Probanden auf die experimentellen Bedingungen (Randomisierung), die unerlässlich ist, um den Erwartungswert der Korrelation zwischen Personmerkmalen und X Null werden zu lassen. Neben der Randomisierung kommen vor allem die Konstanthaltung und die von X unabhängige systematische Variation bekannter Störgrößen in Frage. Durch diese Variation wird eine zweite „unabhängige“ Variable zur Kontrolle eingeführt, und es ist jetzt prüfbar, ob X auf allen Stufen von St denselben Einfluß auf Y ausübt (die statistische Interaktion zwischen X und St ist Null) oder ob dieser Einfluß von St abhängig ist (X und St interagieren). In der experimentellen Psychologie wird, wie gesagt, meistens so verfahren, daß X und St nicht korrelieren. Unterstellt man seinen Daten das lineare Modell einer multiplen Regression (vgl. dazu Schubö et al. in Band 4 dieser Enzyklopädie), so läßt sich das zugrundeliegende Beeinflussungsmodell wie in Abb. 2 darstellen, wobei vorausgesetzt ist, daß X , St und die Interaktionsvariable XSt wechselseitig nicht miteinander korrelieren (diese Variablen sind deshalb in Abb. 2 nicht durch Pfeile verbunden).

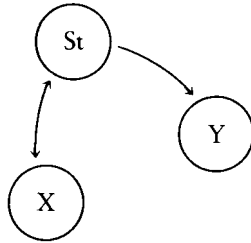


Abb. 1: Der gekrümmte Doppelpfeil weist auf eine korrelative Beziehung zwischen X und St hin, die nicht daraufhin analysiert werden kann, welche Variable unabhängig in bezug auf die andere ist. St beeinflusst Y direkt, während der Zusammenhang zwischen X und Y zum Schein besteht: X und Y korrelieren nur deshalb, weil X mit einer Variablen korreliert, die einen direkten Einfluß auf Y ausübt.

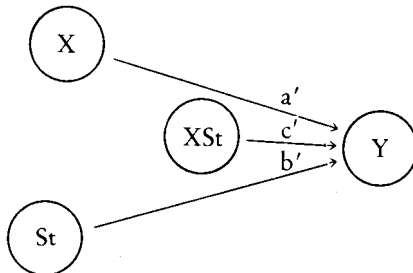


Abb. 2:

In diesem Graphen symbolisieren a' , b' , c' die Einflußgrößen der Variablen X , St und XSt auf Y . Meistens wird mittels der Varianzanalyse geprüft, ob a' , b' und c' Null sind (die Varianzanalyse ist ein Spezialfall der multiplen Regressionsanalyse, und a' , b' , c' sind die (multiplen) Korrelationen zwischen X und Y , St und Y sowie XSt und Y). Ist $c' \neq 0$ und $a' = 0$, spricht man von einer für X disordinalen Interaktion (vgl. Bredenkamp, 1980): Die Richtung der Beeinflussung von X auf Y hängt von dem Wert der Störvariablen ab. Wenn $c' \neq 0$ und $a' \neq 0$, ist für die experimentelle Variable X aus diesen Informationen nicht ableitbar, ob X mit St derart interagiert, daß die gleiche Beeinflussungsrichtung von X auf Y unter allen Werten von St vorliegt (Interaktion für X ist ordinal). Sollte die Interaktion für X disordinal sein, kann nicht von einer Kausalbeziehung zwischen X und Y gesprochen werden. Diese kann nur dann vorliegen, wenn $a' \neq 0$ und $c' = 0$, oder wenn $a' \neq 0$ und $c' \neq 0$, zusätzlich aber gezeigt worden ist, daß die Interaktion an der Beeinflussungsrichtung von X auf Y nichts ändert (vgl. dazu Bredenkamp, 1982).

Mit der Nennung dreier Techniken ist das Reservoir notwendiger Kontrollen, damit das Experiment dem Anspruch als Prüféxperiment von Kausalhypothesen gerecht werden kann, nicht ausgeschöpft. Ausführlich hierüber informieren Hager und Westermann in diesem Band (vgl. auch Bredenkamp, 1980). Später werden wir noch auf einen Aspekt der Kontrolle zu sprechen kommen, welche die Validität des statistischen Schlusses sichern soll.

In Abb. 1 und Abb. 2 sind sog. rekursive Systeme dargestellt, die dadurch ausgezeichnet sind, daß in den Graphen keine Zyklen auftreten. Es gibt keinen Pfad von einer Variablen zu einer anderen Variablen, von der aus man wieder zum Ausgangspunkt zurückkommt. Der Doppelpfeil in Abb. 1 besagt nur, daß die Richtung der Beziehung zwischen X und St nicht analysiert wird. In einem nicht-rekursiven System würde dagegen in Abb. 1 ein Pfeil von X nach St und ein anderer Pfeil von St nach X laufen. Derartige Systeme werden hier nicht betrachtet (vgl. dazu Hummell und Ziegler, 1976). In rekursiven Systemen heißen solche Variablen, von denen nur Pfeile ausgehen, exogen, während Variablen, bei denen wenigstens ein Pfeil ankommt, endogen genannt werden. Üblicherweise sind also die unabhängigen Variablen eines Experiments exogene, die abhängigen Variablen endogene Variablen innerhalb eines rekursiven Systems; allerdings finden sich, wie noch gezeigt wird, auch Beispiele, in denen manche unabhängige Variable endogen ist. Ein rekursives System soll nur dann kausal heißen, wenn es entweder keine Interaktionsvariablen enthält, oder wenn gezeigt werden kann, daß die Interaktionsvariablen die Beeinflussungsrichtung der interessierenden Variablen auf andere nicht modifizieren.

Die Formulierung rekursiver Systeme ist selbstverständlich nicht auf die experimentelle Psychologie beschränkt, sondern auch in der Korrelationsforschung möglich, wobei unter Korrelationsforschung die Analyse korrelativer Bezie-

hungen zwischen Variablen aus einer Untersuchung zu verstehen ist, der nicht die Merkmale „Manipulation“ wenigstens einer Variablen und „Randomisierung“ zukommen (s.o.). Z.B. könnten für die Variablen „Allgemeine Intelligenz“ (AI), „Intelligenz in der Wahrnehmung des Lehrers“ (IW) und „Zeugnisnote“ (ZN) die rekursiven Systeme in Abb. 3a und Abb. 3b formuliert werden (vereinfachtes Beispiel aus Brandstädter und Bernitzke, 1976).

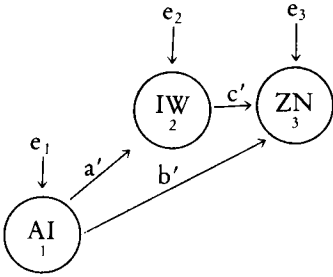


Abb. 3a:

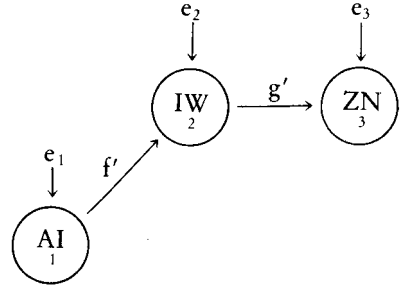


Abb. 3b:

In Abb. 3a wird ein direkter und ein über IW vermittelter Einfluß von AI auf ZN postuliert, während in Abb. 3b nur ein vermittelter Einfluß angenommen wird. Die impliziten Variablen e bezeichnen durch die Problemstellung nicht berücksichtigte Größen, durch deren Einführung man formal eine Schließung des Systems erreicht. Systeme wie in Abb. 3a und Abb. 3b heißen lineare Kausalstrukturen, wenn die Beziehungen zwischen den Variablen ausschließlich linear sind; nur derartige Systeme sollen hier betrachtet werden.

Das System in Abb. 3a heißt vollständig, da Wirkungen zwischen allen Variablenpaaren postuliert werden. Dagegen heißt das System in Abb. 3b unvollständig, da es nicht alle $\binom{m}{2}$ Beziehungen zwischen den m expliziten Variablen enthält.

Von einer kausalen Wirkung von X auf Y innerhalb einer gegebenen Kausalstruktur soll nur dann gesprochen werden, wenn von X auf Y nachweisbar direkte und/oder indirekte Einflüsse bestehen (und wenn die aufgrund der Kausalstruktur vorhergesagten Korrelationen zwischen den Variablen den empirisch ermittelten Korrelationen entsprechen; vgl. Hummell und Ziegler, 1976). Damit sind zwei Prüfbedingungen angesprochen, auf die jetzt eingegangen werden soll. Zuvor sei betont, daß die Aussage „X wirkt kausal auf Y“ immer nur innerhalb eines bestimmten rekursiven Systems gilt oder nicht zutrifft.

Das übliche Experiment versucht durch die Anwendung von Kontrolltechniken zu erreichen, daß der gesamte kausale Einfluß von X auf Y als ausschließ-

lich direkter Einfluß analysiert werden kann. Dieses Vorgehen ist in dem Maße erfolgreich, wie gewährleistet ist, daß durch die Manipulation von X nicht andere Variablen, die auf Y einwirken, verändert werden (interne Validität). Die Prüfung der Aussage „X wirkt kausal auf Y“ geschieht derart, daß statistisch die Hypothese $a'=0$ getestet wird (äquivalent dem Test der Hypothese, daß die Korrelation zwischen X und Y Null ist). Wenn diese Hypothese $a'=0$ abgelehnt werden kann, gilt die Aussage „X wirkt auf Y kausal“ als vorläufig bewährt, wobei noch durch die Einführung von sog. Maßen der praktischen Signifikanz (vgl. dazu Hager und Westermann in diesem Band) zusätzlich gefordert werden kann, daß a, die Schätzung für a' , einen gewissen Wert übersteigt. Vorauszusetzen ist außerdem immer, daß, bezogen auf Abb. 2, $c'=0$ oder die Interaktion zwischen X und St nichts an der Beeinflussungsrichtung von X auf Y unter verschiedenen Werten von St ändert. Wenn jedoch die Hypothese $a'=0$ angenommen werden muß, liegt eine notwendige, aus verschiedenen Gründen jedoch noch nicht hinreichende Bedingung (vgl. dazu Bredenkamp, 1980) für den Schluß or, X wirke nicht kausal auf Y.) Auf statistische Erfordernisse beim Test derartiger Hypothesen kommen wir in Abschnitt 2 zu sprechen. Weitere Prüfungen sind nicht möglich, sofern im rekursiven System nur direkte Wirkungen postuliert werden und keine Vorinformationen über die Größe der Einflüsse a' , b' , c' existieren.

In der Korrelationsforschung lassen sich ebenfalls Kausalhypothesen prüfen. Für die Prüfung von Kausalhypothesen ist nicht die experimentelle Kontrollmöglichkeit entscheidend, sondern die Konzeption eines kausalen rekursiven Systems. Die Kontrolle im Experiment soll nur bewirken, daß allein direkte kausale Einflüsse analysiert werden können. In der Korrelationsforschung dagegen wird wegen eingeschränkter Kontrollmöglichkeiten von vornherein von direkten und indirekten Einflüssen ausgegangen. Die rekursive Pfadanalyse ist das Analyseverfahren, um Kausalhypothesen in der Korrelationsforschung zu überprüfen. Wir gehen hier nur auf die rekursive lineare Pfadanalyse ein (vgl. etwa Brandtstädter und Bernitzke, 1976; Hummell und Ziegler, 1976; Kerlinger und Pedhazur, 1973). Die Anwendung dieses Verfahrens setzt voraus:

- (1) Die Konzeption eines linearen rekursiven Systems.
- (2) Die expliziten Variablen sind metrisch und fehlerfrei gemessen (zur Lockerung dieser Annahmen siehe Hummell und Ziegler, 1976).
- (3) die impliziten Variablen e (vgl. Abb. 3), welche die endogenen Variablen beeinflussen, sind mit keiner anderen Variablen des Systems korreliert.
- (4) Alle Beziehungen zwischen den Variablen sind ausschließlich linear und additiv.
- (5) Interaktionseffekte gibt es nicht.

¹) Diese Darstellungen unterstellen eine multiple Regressionsanalyse experimenteller Daten, die immer dann möglich ist, wenn varianzanalytische Strukturmodelle mit festen Effekten den Daten zugrunde liegen.

Können diese Annahmen für die Beispiele in Abb. 3a und 3b als gültig unterstellt werden, und liegen die Variablen dieser Systeme als Standardwerte Z mit dem Mittelwert 0 und der Standardabweichung 1 vor, läßt sich für die Daten des Modells in Abb. 3a schreiben:

- 1) $Z_1 = e_1$
- 2) $Z_2 = aZ_1 + e_2$
- 3) $Z_3 = bZ_1 + cZ_2 + e_3$,

wobei a , b , c Schätzungen der Parameter a' , b' , c' sind. Multipliziert man Gleichung 2) mit Z_1 , summiert über alle N Probanden und dividiert durch N , erhält man:

$$r_{12} = a.$$

Die gleiche Verfahrensweise führt zu:

$$\begin{aligned} r_{13} &= b + cr_{12} \\ r_{23} &= br_{12} + c, \end{aligned}$$

wobei die beiden letzten Gleichungen zur Schätzung der Pfadkoeffizienten b und c verwendet werden:

$$b = \frac{r_{13} - r_{12}r_{23}}{1 - r_{12}^2} \quad c = \frac{r_{23} - r_{12}r_{13}}{1 - r_{12}^2}.$$

Das in Abb. 3a veranschaulichte System ist vollständig, und wie gezeigt worden ist, werden alle möglichen Korrelationen benötigt, um die Einflußgrößen (Pfadkoeffizienten) zu schätzen. Sofern also keine Vorinformationen über diese Größen vorliegen, lassen sich keine Korrelationen zwischen den Variablen eines vollständigen Systems prognostizieren. Dies trifft auch für das Experiment zu, soweit es bisher besprochen wurde. Im Unterschied zum Experiment wird aber in einer vollständigen linearen Kausalstruktur die kausale Wirkung auf direkte und indirekte Einflüsse zurückgeführt. Dies wird sichtbar, wenn in den Gleichungen für r_{13} und r_{23} a für r_{12} eingesetzt wird:

$$\begin{aligned} r_{13} &= b + ac \\ r_{23} &= c + ab. \end{aligned}$$

r_{13} ist also eine Schätzung des kausalen Einflusses von Z_1 auf Z_3 innerhalb der linearen Kausalstruktur in Abb. 3a, der sich aus einer direkten Wirkung b und einer indirekten Wirkung ac zusammensetzt. Auch Z_2 hat einen direkten (c) Einfluß auf Z_3 , aber keinen indirekten, da von Z_2 kein weiterer Pfad nach Z_3 führt: ab gibt hier die in r_{23} enthaltene Scheinbeziehung an.

Vollständige Modelle nun lassen sich, genau wie dem Experiment zugrunde liegende Modelle, nur derart prüfen, daß für die Pfadkoeffizienten getestet

wird, ob sie statistisch gesehen von Null abweichen. Ergibt sich, daß alle Pfadkoeffizienten von Null abweichen, ist dies noch keine Bestätigung für das geprüfte vollständige Modell. Jede andere vollständige lineare Kausalstruktur mit den gleichen expliziten Variablen würde gleichermaßen gut für die Daten passen. Wenn jedoch zusätzlich eine zeitliche Ordnung zwischen den Variablen postuliert werden kann, derart z.B., daß wie in Abb. 3a AI vor IW und IW vor ZN steht, könnte das Modell als bewährt angesehen werden. Ist jedoch wenigstens einer der Pfadkoeffizienten Null, kann ein unvollständiges Modell formuliert werden, in dem nur einige Korrelationen zur Schätzung der Pfadkoeffizienten benötigt werden. In diesem Fall lassen sich andere Korrelationen prognostizieren, und die Prognosen können mit den tatsächlichen erhaltenen Daten verglichen werden.

Wenn die empirisch erhaltenen Korrelationen zu Abb. 3a bei $N = 100$ Probanden $r_{12} = 0.40$, $r_{13} = 0.25$ und $r_{23} = 0.60$ betragen, läßt sich ermitteln:

$$a = 0.40 \quad b = 1/84 \quad c = 50/84.$$

Da b nahezu Null ist, wurde das Modell aus Abb. 3b formuliert, dessen Strukturgleichungen für Stichprobendaten lauten:

$$\begin{aligned} 4) \quad Z_1 &= e_1 \\ 5) \quad Z_2 &= fZ_1 + e_2 \\ 6) \quad Z_3 &= gZ_2 + e_3. \end{aligned}$$

Durch Multiplikation von Gleichung 5) mit Z_1 , Aufsummierung und Division durch N ergibt sich:

$$r_{12} = f.$$

Ein entsprechendes Vorgehen führt zu:

$$\begin{aligned} r_{13} &= gr_{12} = fg. \\ r_{23} &= g. \end{aligned}$$

g kann also auf zweierlei Weise geschätzt werden. Aus $g = r_{23}$ ergibt sich die Prognose für eine Korrelation, die zur Schätzung der Pfadkoeffizienten nicht benötigt wird:

$$r_{13}^* = r_{12}r_{23} = 0.24.$$

Verglichen mit dem tatsächlichen Wert $r_{13} = 0.25$ ist die Prognose recht genau, und man kann auf einen Signifikanztest verzichten. Ist die Übereinstimmung nicht derart deutlich, muß aus $r_{13}^* = r_{12}r_{23}$ eine Prognose abgeleitet werden, die sich statistisch prüfen läßt. Würde man folgern, daß die Partialkorrelation

$$r_{13.2} = \frac{r_{13} - r_{12}r_{23}}{\sqrt{1-r_{12}^2} \sqrt{1-r_{23}^2}}$$

Null sein muß, so ginge in diese Folgerung die modellunverträgliche Annahme ein, daß Z_2 auf Z_1 und Z_3 einwirkt (vgl. Abb. 3b). Folgert man jedoch, daß die semipartielle Korrelation

$$r_{1(3.2)} = \frac{r_{13} - r_{12}r_{23}}{\sqrt{1 - r_{23}^2}}$$

Null ist, so liegt eine modellverträgliche Prognose vor, in die die Annahme eingeht, daß Z_2 sich auf Z_3 auswirkt. Diese Prognose läßt sich statistisch prüfen.

Nummehr soll die rekursive Pfadanalyse auf ein komplexeres Beispiel angewendet werden. Dieses Beispiel ist der experimentellen Psychologie entnommen (ein komplexes Beispiel aus der Korrelationsforschung findet sich bei Brandstädter, 1976). Auch im Experiment ist es nicht immer möglich, die verschiedenen Variablen eines rekursiven Systems unabhängig voneinander so zu variieren, daß nur direkte kausale Einflüsse analysiert werden können. Will man z. B. den Einfluß der Bildhaftigkeit des Lernmaterials auf die Gedächtnisleistung prüfen, so sind andere Variablen mit der Bildhaftigkeit konfundiert. So entsteht etwa beim Lernen von Sätzen das Problem, ob die Bildhaftigkeit oder die Verständlichkeit der Sätze die Gedächtnisleistung determiniert. Wipich und Bredenkamp (1979) haben argumentiert, daß die Bildhaftigkeit der Sätze die Variable „Verständlichkeit“ beeinflusst und nicht umgekehrt. Nimmt man die Annahmen hinzu, daß die Bildhaftigkeit der Substantive in Subjekt- und Objektposition die Bildhaftigkeit des Satzes determinieren, so läßt sich das Modell in Abb. 4 formulieren:

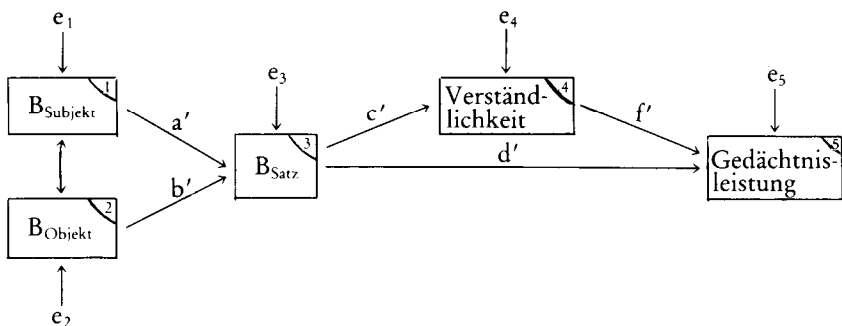


Abb. 4:

Die Gedächtnisleistung wurde unter fünf verschiedenen experimentellen Bedingungen erhoben, die erst später beschrieben werden sollen. Entsprechend wird diese Variable in Tab. 1, die die empirisch erhaltenen Korrelationen wiedergibt, mit 5a bis 5e bezeichnet. Folgende Strukturgleichungen lassen sich nach dem Modell in Abb. 4 für die Daten schreiben:

- 7) $Z_1 = e_1$
- 8) $Z_2 = e_2$
- 9) $Z_3 = aZ_1 + bZ_2 + e_3$
- 10) $Z_4 = cZ_3 + e_4$
- 11) $Z_5 = dZ_3 + fZ_4 + e_5$

Tabelle 1:

	1	2	3	4
1				
2	0.93			
3	0.96	0.96		
4	0.90	0.90	0.93	
5a	0.78	0.74	0.73	0.75
5b	0.82	0.78	0.80	0.74
5c	0.91	0.92	0.93	0.87
5d	0.21	0.21	0.22	0.26
5e	0.86	0.87	0.88	0.83

Durch Multiplikation der Gleichung 9) mit Z_1 (bzw. Z_2), Aufsummierung und Division durch N erhält man:

$$\begin{aligned} r_{13} &= a + br_{12} \\ r_{23} &= ar_{12} + b \end{aligned}$$

Diese beiden Gleichungen genügen, um $a = 0.50$ und $b = 0.50$ zu schätzen.²⁾ Weiterhin läßt sich zeigen, daß

$$r_{34} = c = 0.93.$$

Da r_{14} und r_{24} für keine Schätzung benötigt werden, lassen sie sich prognostizieren:

$$\begin{aligned} r_{14} &= cr_{13} = ac + bcr_{12} = 0.89 \text{ (tatsächlicher Wert: 0.90)} \\ r_{24} &= cr_{23} = acr_{12} + bc = 0.89 \text{ (tatsächlicher Wert: 0.90)}. \end{aligned}$$

Weiterhin läßt sich zeigen, daß:

$$\begin{aligned} r_{15} &= dr_{13} + fr_{14} = ad + bdr_{12} + acf + bcf r_{12} \\ r_{25} &= dr_{23} + fr_{24} = adr_{12} + bd + acfr_{12} + bcf \\ r_{35} &= d + fr_{34} = d + cf \\ r_{45} &= dr_{34} + f = cd + f. \end{aligned}$$

Für zwei unbekannte Größen d und f stehen vier Gleichungen zur Verfügung; zwei werden jedoch nur benötigt. Entscheidet man sich dafür, die Gedächtnisleistung aus der Bildhaftigkeit des Subjekts und Objekts (Z_1, Z_2) prognostizieren zu wollen, schätzt man aus den Gleichungen für r_{35} und r_{45} die Größen d und f . Tab. 2 enthält die Werte für d , f , r_{15}^* und r_{25}^* . Beim Vergleich mit den tatsächlich erhaltenen Korrelationen aus Tab. 1 fällt die gute Übereinstimmung auf. Ferner fällt auf, daß vermutlich nur für Variable 5a das vollständige Modell aus Abb. 4 benötigt wird, während für die Variablen 5b, 5c und 5e das Modell in Abb. 5a und für Variable 5d das Modell in Abb. 5b ausreicht.

Tabelle 2:

	d	f	r_{15}	r_{25}
Variable 5a	0.24	0.53	0.70	0.70
5b	0.83	-0.03	0.77	0.77
5c	0.89	0.04	0.89	0.89
5d	-0.16	0.41	0.21	0.21
5e	0.80	0.09	0.85	0.85

Aus den Strukturgleichungen für Modell 5a lassen sich dieselben Pfadkoeffizienten a , b , c wie für das Modell 4 gewinnen, außerdem ist jetzt $d = r_{35}$. Nunmehr läßt sich prognostizieren für Variable 5:

$$\begin{aligned} r_{15}^* &= dr_{13} = ad + bdr_{12} \\ r_{25}^* &= dr_{23} = adr_{12} + bd \\ r_{45}^* &= dr_{34} = cd. \end{aligned}$$

*) Bei der Schätzung der Pfadkoeffizienten können sich erhebliche Fehler einstellen, wenn die Variablen so hoch miteinander korrelieren, wie das in diesem Beispiel der Fall ist. Um derartige Schätzfehler zu vermeiden, sollte man deshalb mit einem Modell arbeiten, das nur die Variablen Z_1, Z_2 und Z_5 erhält. Nur zur Demonstration eines komplexeren Modelies wird hier auf diese empfehlenswerte Reduktion verzichtet.

Grundsätzlich ist zur Prognose der Korrelationen folgendes zu sagen: r_{ij} , mit $i < j$ wird immer gewonnen, indem die Gleichung für Z_i mit Z_j multipliziert wird. Die Prognose erfolgt durch Verknüpfung der Pfadkoeffizienten und nicht analysierter Korrelationen (in diesem Beispiel r_{12}).

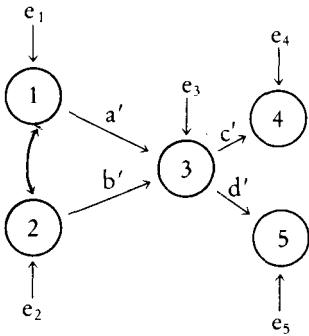


Abb. 5a:

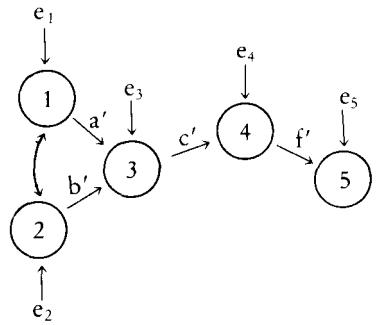


Abb. 5b:

Dieses Modell, das eine Beziehung weniger als Modell 4 postuliert, ermöglicht auch eine Prognose mehr. Es ergeben sich, wie ein Vergleich der Tab. 3 mit Tab. 1 zeigt, gute Übereinstimmungen.

Tabelle 3:

		r_{15}	r_{25}	r_{35}
Variable	5b	0.77	0.77	0.74
	5c	0.89	0.89	0.86
	5e	0.84	0.84	0.82

Aus den Strukturgleichungen für Modell 5b lassen sich wiederum dieselben Pfadkoeffizienten a, b, c schätzen, und f ergibt sich zu r_{45} . Prognostizieren lassen sich für Variable 5:

$$\begin{aligned} r_{15}^* &= fr_{14} = fcr_{13} = acf + bcfr_{12} \\ r_{25}^* &= fr_{24} = fcr_{23} = acfr_{12} + bcf \\ r_{35}^* &= fr_{34} = fc. \end{aligned}$$

Die Prognosen und tatsächlichen Werte für Variable 5d enthält Tab. 4.

Tabelle 4:

$r_{15}^* = 0.23$	$r_{25}^* = 0.23$	$r_{35}^* = 0.24$
$r_{15} = 0.21$	$r_{25} = 0.21$	$r_{35} = 0.22$

Kommen wir nun auf die Interpretation zu sprechen. Als zentral soll die Aufklärung des Zusammenhangs zwischen der Bildhaftigkeit der Sätze und der

Gedächtnisleistung angesehen werden. Das Modell in Abb. 4 postuliert einen direkten Einfluß, dessen Größe durch den Pfadkoeffizienten d geschätzt wird, sowie einen indirekten Einfluß cf . Dieses Modell wurde für Variable 5a akzeptiert, wobei $d = 0.24$ und $cf = 0.49$. Bei Variable 5a handelt es sich um die Reproduktionsleistung nach intentionalem Lernen: Die Vpn wußten, daß die Sätze reproduziert werden mußten. In dieser Situation besteht also ein direkter und ein über die Verständlichkeit vermittelter Einfluß der Bildhaftigkeit der Sätze auf die Reproduktionsleistung. Für die Variablen 5b, 5c und 5e wurde das Modell in Abb. 5a konzipiert, nachdem sich bei der Berechnung der Pfadkoeffizienten für das komplexere Modell in Abb. 4 herausgestellt hatte, daß f nahe Null liegt; auch die Korrelationen ließen sich mit Hilfe dieses Modells gut reproduzieren. Wenn man will, kann man mit Hilfe statistischer Hypothesentests den Vergleich beider Modelle vornehmen. Da im Modell der Abb. 4 die Variable „Gedächtnisleistung“ durch die Prädiktoren Z_3 und Z_4 , im Modell der Abb. 5a jedoch nur durch den Prädiktor Z_3 determiniert wird, sollte statistisch gesehen Z_4 keine zusätzliche Varianzaufklärung leisten: $R_{5,3,4}^2 - r_{53}^2$ sollte nicht signifikant von Null abweichen. Dieser Betrag gibt die Größe der quadrierten semipartiellen Korrelation zwischen Gedächtnisleistung und Verständlichkeit wieder, nachdem aus dieser Variablen der Einfluß der Bildhaftigkeit der Sätze ausgeschaltet wurde. Für die Variablen 5b, 5c, 5e ist dieser Betrag mit 0.01, 0.01 und 0.01 sehr klein und nicht signifikant. Die Tests, die $R_{5,3,4}^2 - r_{53}^2$ auf Signifikanz prüfen, sind Prüfungen der Bedeutsamkeit des Pfadkoeffizienten f' aus Modell 4. Das Modell in Abb. 5a postuliert nur einen direkten Einfluß der Bildhaftigkeit der Sätze auf die Gedächtnisleistung. Bei den Variablen 5b und 5c handelt es sich ebenfalls um Gedächtnisleistungen nach dem intentionalen Lernen. Allerdings hatten die Pbn Zusatzaufgaben während des Lernens zu verrichten, die nach dem levels of processing-Ansatz verschiedene Verarbeitungsformen induzieren sollten: Für Variable 5b sollte die Bildhaftigkeit eine geringere Rolle als für Variable 5c spielen, was auch, wenn man die beiden Korrelationen r_{35} vergleicht, der Fall war. Überraschend ist allerdings der Befund, daß auch für Variable 5b nur ein direkter Einfluß der Bildhaftigkeit besteht. Verglichen mit Variable 5a führt also beim intentionalen Lernen jede der verwendeten Zusatzaufgaben zu einem direkten Einfluß der Bildhaftigkeit; nur ohne Zusatzaufgabe besteht auch ein indirekter Einfluß. Nach inzidentellem Lernen (Variablen 5d und 5e) besteht nur nach „kognitiv tiefer“ Verarbeitung (Variable 5e) ein direkter Effekt der Bildhaftigkeit, während nach „oberflächlicher“ Verarbeitung (Variable 5d) allein ein indirekter Effekt besteht. Die Angemessenheit des Modells in Abb. 5b für Variable 5d kann dadurch getestet werden, daß wegen $r_{35}^* = cf = r_{34} \cdot r_{45}$ folgt: Die semipartielle Korrelation $r_{3(5,4)}$ weicht nicht signifikant von Null ab. Der Wert für diese Korrelation beträgt für Variable 5d -0.02 und ist insignifikant.

Die vorgetragenen Ergebnisse haben u.E. bisher nicht gesehene Konsequenzen beim Versuch der Zusammenführung des levels of processing-Ansatzes

mit der „imagery“-Theorie Paivios, auf die hier nicht einzugehen ist. Einige dieser Ergebnisse kamen unerwartet; ihre Interpretation ist im Rahmen der behandelten Modelle gültig. Die Richtigkeit derartiger Modelle läßt sich selbstverständlich nicht beweisen. Es sind immer andere Modelle denkbar, die ebenfalls gut für die Daten passen. Die Pfadanalyse ist, wie Kerlinger und Pedhazur (1973) zu Recht feststellen, ein Verfahren, das die Zurückweisung unhaltbarer Kausalmodelle eher als die Bestätigung eines von verschiedenen rivalisierenden Kausalmodellen leistet. Die Pfadanalyse ist auch kein Verfahren, das zu Modellen führt! Die Theorie- oder Modellbildung wird dem Forscher nicht abgenommen. Sie ist unabdingbare Voraussetzung für die Anwendung der Pfadanalyse, die ein Verfahren zur Prüfung von Theorien ist und möglicherweise Hinweise, wie gezeigt wurde, dafür gibt, an welchen Stellen eine Theorie zu modifizieren ist. Leider sind die Möglichkeiten der Pfadanalyse zur Prüfung kausaler Hypothesen in der Korrelationsforschung bisher viel zu selten genutzt worden. Einen methodisch völlig anders gearteten Ansatz zur Analyse von Kausalbeziehungen beschreibt Lehmann (1980).

Das behandelte Beispiel aus der experimentellen Psychologie zeigt Analysemöglichkeiten auf, wenn unabhängige Variablen konfundiert sind: Wenigstens eine dieser Variablen wird als endogen konzipiert. Dieser Fall liegt bei der Anwendung der Kovarianzanalyse generell vor: Im Rahmen eines rekursiven linearen Modells betrachtet wird angenommen, daß die Kovariate X sich auf die endogene unabhängige Variable U und die abhängige Variable Y auswirkt (vgl. Abb. 6a). Die Kovarianzanalyse prüft, ob die Einflußgröße c' Null ist, indem die semipartielle Korrelation zwischen Y und U , aus der der Einfluß der Kovariaten eliminiert wird, auf Signifikanz getestet wird. Dieses Vorgehen ist sinnvoll, da U nur einen direkten kausalen Einfluß auf Y ausübt. In der Korrelation zwischen U und Y ist auch noch eine Scheinbeziehung zwischen beiden Variablen enthalten ($a'b'$), und deshalb interessiert allein die Größe von c , die den gesamten kausalen Einfluß von U auf Y erfaßt (übrigens läßt man in der experimentellen Psychologie die Scheinbeziehung durch zufällige Zuweisung der Probanden auf die experimentellen Bedingungen von vornherein Null werden, da in diesem Fall $a'=0$).

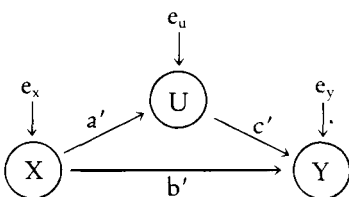


Abb. 6a: Modell für die Kovarianzanalyse.

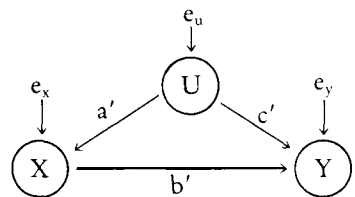


Abb. 6b: Unzutreffendes Modell für die Kovarianzanalyse.

Nun finden sich häufig Beispiele für die Verwendung einer Kovarianzanalyse in der Situation, die in Abb. 6b dargestellt ist. In diesem Fall wirkt sich die unabhängige Variable auf die Kovariante aus, und man möchte wissen, ob ein Einfluß von U auf Y besteht, nachdem die Unterschiede in der Kovariaten statistisch ausgeglichen wurden. Die Anwendung der Kovarianzanalyse ist in diesem Fall nicht gerechtfertigt. r_{uy} gibt in Abb. 6b den gesamten Einfluß von U auf Y wieder. Will man prüfen, ob ein direkter Einfluß von U auf Y besteht, testet man die semipertielle Korrelation $r_{U(Y.X)}$ auf Signifikanz.

2. Zum Problem der Validität des statistischen Schlusses

Kausalmodelle werden in der experimentellen Psychologie und der Korrelationsforschung statistisch überprüft. Diese Feststellung führt zur Frage der Validität des statistischen Schlusses. In der experimentellen Psychologie implizieren psychologische Hypothesen häufig die zur Nullhypothese eines statistischen Tests alternative Hypothese, so daß auf das Zutreffen der Nullhypothese erkannt werden können muß, damit die psychologische Hypothese falsifizierbar ist. Dies ist nur dann möglich, wenn die Wahrscheinlichkeit des β -Fehlers ebenso wie die des α -Fehlers kontrolliert wird. Es wird also empfohlen, eine minimale Effektgröße festzulegen, die mit der Wahrscheinlichkeit $1 - \beta$ entdeckt werden soll, und α und β als kleine Beträge festzulegen (Bredenkamp, 1969, 1972). Für verschiedene statistische Verfahren läßt sich dann der benötigte Stichprobenumfang bestimmen. Die bei Cohen (1977) publizierten Tabellen erleichtern diese Bestimmung wesentlich. Das hier nur andeutungsweise vorgestellte Verfahren (ausführlich dazu Bredenkamp, 1980; Hager und Westermann in diesem Band) ist auch dann notwendig, wenn die statistische Nullhypothese durch die psychologische Hypothese impliziert wird, um zu viele unverdiente Bestätigungen der psychologischen Hypothesen zu vermeiden. Außerdem ist eine derartige Planung auch für die pfadanalytische Überprüfung rekursiver linearer Systeme vonnöten, wenn auch hier statistische Tests zur Prüfung des Modells durchgeführt werden und man sich nicht mit der augenscheinlichen Übereinstimmung der reproduzierten mit den tatsächlich erhaltenen Korrelationen begnügt. Schwierigkeiten bereitet dieses Verfahren insofern, als zu Beginn eines Forschungsprogramms die Festlegung einer minimalen Effektstärke unmöglich erscheint. Diesem Problem ist man jedoch nicht entronnen, wenn man auf eine derartige Festlegung verzichtet, da man durch die unbegründete Wahl des Stichprobenumfangs auf eine minimale Effektstärke festgelegt worden ist, die mit vorgegebenen Wahrscheinlichkeiten α und $1 - \beta$ entdeckt werden kann. In diesem Fall bleibt auch die Frage, ob Versuchsergebnisse durch andere Forscher repliziert worden sind, unentscheidbar (vgl. Bredenkamp, 1980). Wenn jedoch anerkannt wird, daß psychologische Hypothesen oder Kausalmodelle nicht aufgrund einzelner Untersuchungen, sondern ganzer Forschungsprogramme falsifiziert werden kön-

nen, läßt sich das Problem der Festlegung der Effektgröße entschärfen. Legt man diese „willkürlich“ auf einen bestimmten Wert wie z.B. $R^2 = 0.10$ fest, so hat sich eine psychologische Hypothese, die die statistische Alternativhypothese impliziert, in einem Forschungsprogramm dann bewährt, wenn $Q = \alpha^x (1 - \alpha)^{m-x} / (1 - \beta)^x \beta^{m-x} < 1$, wobei m die Anzahl der Untersuchungen und x die Anzahl signifikanter Resultate ist. Aber auch bei $Q > 1$ könnte behauptet werden, daß die psychologische Hypothese gestützt ist; da von einem Minimaleffekt $R^2 = 0.10$ ausgegangen wurde, der sich als zu groß herausgestellt habe, müsse β tatsächlich viel größer als festgelegt gewesen sein, so daß etwa 3 von 10 signifikanten Ergebnissen die Hypothese eines kleineren Effekts stützen. Eine derartige Behauptung müßte begründet werden können (z.B. zu geringe Durchschlagskraft der unabhängigen Variablen), und nur im Falle einer Begründung (und nicht einer beliebigen Exhaustion „negativer“ Befunde) wäre eine prüfbare Hypothese generiert worden. An dieser Stelle zeigt sich, daß mit der skizzierten Vorgehensweise Falsifikationen logisch und empirisch nicht erzwungen werden können. Der Forscher kann sich aufgrund in diesem Sinne geplanter Untersuchungen nur zur Falsifikation und zum Neuentwurf von Hypothesen *entschließen*. Die Planung der Untersuchungen ist so anzulegen, daß ein derartiger Entschluß überhaupt ermöglicht wird. Eine Planung, wie sie hier skizziert wurde (ausführlich dazu Bredenkamp, 1980), würde vermutlich auch erreichen, daß sowohl seitens des Forschers wie der Herausgeber von Fachzeitschriften „negative“ Resultate veröffentlicht werden, so daß die häufig beklagte Kumulation des statistischen α -Fehlers in den Publikationen vermieden werden könnte.

Gegen die skizzierte Vorgehensweise sind durch Deppe (1977) und Glaser (1979) Bedenken angemeldet worden. Deppe (1977) weist, ganz im Sinne der obigen Ausführungen, darauf hin, daß ein Modell durch das Ergebnis eines Signifikanztests logisch nicht zu widerlegen ist. Andererseits schreibt er: „Wenn in *vielen* Experimenten zum Begriffserwerb z.B. ein negativ beschleunigter Abfall der Fehlerkurven beobachtet wird, und wenn ein Modell hier einen konstanten Abfall der Fehler voraussagt, ist eher anzunehmen, daß das Modell systematisch falsche Annahmen macht, als wenn die Fehlerkurve nur einmal beobachtet wurde“ (Deppe, 1977, 166). über derartige Abweichungen der tatsächlichen von den prognostizierten Resultaten kann man mit Hilfe eines Signifikanztests entscheiden, und wenn aus mehreren Untersuchungen die dem Modell widersprechenden Ergebnisse der Tests konvergieren, kann man sich zu einer Falsifikation und Abänderung des Modells entschließen, vorausgesetzt, die Tests wurden im Sinne obiger Ausführungen geplant. Die Ausführungen Deppes scheinen uns der dargestellten Konzeption des Hypothesentests nicht zu widersprechen, wenn anerkannt wird, daß die Falsifikation nur aufgrund eines methodologischen Beschlusses möglich ist, z.B.: „Akzeptiere bei kleinen Fehlerwahrscheinlichkeiten H_0 als eine der psychologischen Hypothese widersprechende Populationsaussage“ (vgl. Bredenkamp, 1980).

Glaser (1979, 135) anerkennt das skizzierte Verfahren als sinnvoll, betont jedoch, daß es nicht der Falsifikation von psychologischen Hypothesen dienen könne. Nur deterministisch formulierte Hypothesen sind falsifizierbar, nicht jedoch statistische Hypothesen, da sie kein Ereignis verbieten. Diese Feststellungen treffen selbstverständlich zu und stimmen mit der Aussage Deppes überein, daß das Ergebnis eines Signifikanztests logisch ein Modell nicht widerlegen könne. Glaser (1979) scheint zwar die Möglichkeit zu akzeptieren, aufgrund eines statistischen Tests eine vorgeordnete psychologische Hypothese zu verwerfen - von Falsifikation sollte s.E. nicht gesprochen werden -, wähnt sich aber auf sichererem Boden, wenn psychologische Gesetze von vornherein als Wahrscheinlichkeitsaussagen formuliert werden, die nicht falsifiziert werden können. Dem ist jedoch entgegenzuhalten, daß die de facto bei der Durchführung statistischer Tests geprüften statistischen Hypothesen selten mit den psychologischen identisch sind. Beispiele für diese Behauptung finden sich etwa bei Bredenkamp (1972, 1980) und bei Hager und Westermann in diesem Band. Psychologische Hypothesen beziehen sich auf individuelles Geschehen; die in der empirischen Psychologie üblichen statistischen Tests prüfen aber Populationsaussagen über die Gleichheit oder Verschiedenheit von Parametern. Will man diese Prüfungen nicht von vornherein als sinnlos betrachten, muß ein Bezug zwischen psychologischen Hypothesen und Populationsaussagen hergestellt werden. Deterministisch formulierte psychologische Hypothesen implizieren Populationsaussagen. Auf erstere richtet sich der Falsifikationsanspruch. Falsifikationsinstanzen sind statistische Populationsaussagen, über deren Zutreffen entschieden werden muß. Dabei ist zu gewährleisten, daß die mit dieser Entscheidung verbundenen Fehlermöglichkeiten gering sind, und genau dazu dient das skizzierte Verfahren. Wenn Glaser (1979, S. 125) sagt, daß „eine deterministische psychologische Hypothese . . . mit einem widersprechenden Datum von einer Person, die dem Individuumbereich der Hypothese angehört, erledigt“ ist, scheint er vorauszusetzen, daß immer für einzelne Personen singuläre Existenzsätze formuliert werden können, die im Einklang oder im Widerspruch zur Hypothese stehen. Dies aber ist in der Psychologie eben häufig nicht der Fall (vgl. Bredenkamp 1972, 1980), und dennoch sind die zu prüfenden Hypothesen vielfach deterministisch formuliert. Die implikative Verknüpfung zwischen psychologischer Hypothese und Populationsaussage und daraus resultierende Veränderungen am in der Psychologie üblichen statistischen Test scheinen uns der einzige Weg zu sein, dieser Situation gerecht zu werden. Freilich können dann nur bei Vereinbarung methodologischer Regeln Hypothesen falsifiziert werden. Dies aber ist keine Besonderheit des Forschungsprozesses, die erst durch das statistische Hypothesentesten ins Spiel kommt. Popper (1966) hat in seiner „Logik der Forschung“ immer wieder darauf hingewiesen, daß auch singuläre Existenzsätze Dispositionsprädikate enthalten, die nicht vollständig auf beobachtbare Gegebenheiten zurückgeführt werden können. Die Basissätze müssen deshalb innerhalb der Falsifikationstheorie Poppers (1966) durch Festsetzung aner-

kannt werden. Übernehmen statistische Populationsaussagen die Funktion von Basissätzen, so müssen methodologische Festsetzungen vereinbart werden, die ihre Anerkennung ermöglichen. Die von Glaser (1979) akzeptierte Modifikation des statistischen Hypothesentestens scheint uns in der Psychologie überhaupt nur unter dieser Zielsetzung begründbar zu sein. Bei der pfadanalytischen Überprüfung linearer Kausalstrukturen verhält es sich nicht anders. Auch diese Modelle sind im Grunde deterministisch formuliert. Durch die Hereinnahme der impliziten Variablen wird behauptet, daß die Variation der endogenen Variablen völlig aufgeklärt werden kann. Da diese Variablen unbekannt sind, lassen sich aus dem Modell jedoch nur Korrelationen zwischen den expliziten Variablen ableiten (und nicht etwa linear funktionale Beziehungen), und aufgrund dieser Korrelationen wird das Modell überprüft.

Es sei hier noch kurz auf eine erweiterte Theorie des statistischen Hypothesentestens eingegangen, die Witte (1980) formuliert hat. Danach vollzieht sich die Beurteilung von statistischen Hypothesen in folgenden Schritten:

- (1) α und β werden klein gewählt; es wird eine Effektgröße, die mit der Wahrscheinlichkeit $1 - \beta$ entdeckt werden soll, festgelegt, und der benötigte Stichprobenumfang wird bestimmt. Dieses Vorgehen entspricht den bisherigen Ausführungen.
- (2) Anstelle eines Signifikanztests wird ein likelihood-Test (vgl. dazu Wendt in diesem Band) für zwei einfache Hypothesen durchgeführt, dem der Vorzug gegenüber dem Signifikanztest gegeben wird, weil er nur das eingetretene Ergebnis (etwa $\bar{X}=2$) und nicht auch größere Abweichungen vom Parameter $\mu_0=0$ unter H_0 verarbeitet. Entspricht die Effektstärke etwa einem Parameter $\mu_1=5$, so lautet der Test:

$$\psi = \frac{L(\mu_0=0/\bar{X}=2)}{L(\mu_1=5/\bar{X}=2)}$$

Der Wert für ψ muß die Grenze $g_1 = \frac{1-\alpha}{\beta}$ überschreiten, wenn H_0 , oder $g_2 = \frac{\alpha}{1-\beta}$ unterschreiten, wenn H_1 besser gestützt ist, wobei diese Grenzen der Theorie sequentieller Verfahren von Wald (vgl. dazu Wendt in diesem Band) entnommen wurden. Der Prüfschritt ist nur dann befriedigend ausgefallen, wenn g_1 über- oder g_2 unterschritten wird.

- (3) Sind die beiden ersten Prüfschritte positiv ausgegangen, dann wird ermittelt, ob die größere Stützung der einen Hypothese auf einem hohen likelihood-Wert der akzeptierten oder auf einem geringen likelihood-Wert der abgelehnten Hypothese beruht, indem die likelihoods an der maximalen likelihood gemessen werden. Dabei muß der kritische Quotient $Q_C = 1 - \sqrt{\alpha(1-\beta)}$ durch die in Schritt 2 akzeptierte Hypothese erreicht oder überschritten werden.
- (4) Schließlich ist die Größe des Effekts zu schätzen, wobei Witte (1980) als

Faustregel angibt, daß nicht weniger als 10% der totalen Varianz aufgeklärt sein sollten.

Die besser gestützte Hypothese wird akzeptiert, wenn alle Prüfschritte positiv ausgefallen sind; nur dann wird auch die weniger gestützte Hypothese abgelehnt. Falls wenigstens einer der vier Prüfschritte negativ ausfällt, wird keine Entscheidung getroffen.

Kritisch ist zu diesem Vorgehen anzumerken, daß die Prüfschritte 2 und 3 negativ ausgehen können, wenn ein sehr großer Effekt besteht, der vierte Prüfschritt also positiv ausgefallen ist. Dies wäre etwa der Fall, wenn für normal verteilte Daten die Hypothesen $\mu_0=0$ und $\mu_1=1$ bei $\sigma=1$ getestet werden, \bar{X} aber 10 oder größer ist. Diese Schwierigkeit hängt damit zusammen, daß nur einfache statistische Hypothesen getestet werden. Selten jedoch liegt in der Psychologie dieser Fall vor.

Bei den besprochenen Konzeptionen fand das Bayessche Hypothesentesten keine Berücksichtigung, das Wendt in diesem Band ausführlich darstellt und favorisiert (zur Kritik an der Verwendung des Bayesschen Theorems beim statistischen *Hypothesentesten*, vgl. Rützel, 1979, 1980). Die von Vertretern der Bayes-Statistik oft behauptete und auch aufgezeigte Voreingenommenheit von Signifikanztests gegen die Nullhypothese kann, wenn man einmal die Prämisse der Bayesianer, daß Verteilungsparameter eine Zufallsvariable sind, der Wahrscheinlichkeiten (oder Wahrscheinlichkeitsdichten) zuzuordnen sind, akzeptiert, praktisch durch die eingangs beschriebene Kontrolle der Wahrscheinlichkeiten α und β und deren Identifikation mit kleinen Werten aufgehoben werden (vgl. dazu Bredenkamp, 1972). Allerdings kann man das Bayessche Theorem dann gut verwenden, wenn Hypothesen im Lichte ganzer Forschungsprogramme beurteilt werden und ermittelt werden soll, welche der statistischen Hypothesen besser gestützt ist. Unterteilt man den Parameterraum in zwei Teilklassen H_0 und H_1 , die zu Beginn eines Forschungsprogramms beide für gleich wahrscheinlich gehalten werden, und zerlegt man die Menge aller möglichen Ergebnisse ebenfalls in zwei Teilklassen „signifikant“ und „insignifikant“, so ist bei Verwendung des Bayesschen Theorems immer die Wahrscheinlichkeit für H_0 im Lichte der Daten ($P(H_0/D)$) größer als die für H_1 , wenn der zuvor eingeführte Wert $Q = \alpha^x(1-\alpha)^{m-x}/(1-\beta)^x\beta^{m-x}$ größer als 1 ist, da $P(H_0/D)=Q/(Q+1)$. Unter den genannten Voraussetzungen ist Q ein Bayessches Stützmaß; hinzu kommt, daß α und β , wie besprochen, kontrolliert wurden.

3. *Dynamische Modelle*

Die Zeit und damit korrelierte Veränderungen der Variablen wurden bisher nicht berücksichtigt. Die in Abschnitt 1 besprochenen Modelle eignen sich

nach Hummell und Ziegler (1976) dann zur Darstellung kausaler Prozesse, wenn eine der folgenden Bedingungen erfüllt ist:

- (1) Auf eine Änderung der exogenen Faktoren erfolgt sofort eine Änderung der endogenen Variablen.
- (2) Zum Zeitpunkt der Beobachtung befindet sich der Prozeß in einem Gleichgewichtszustand: Die Werte aller Variablen ändern sich nicht mehr.
- (3) Es liegt ein Prozeß zugrunde, der die Korrelationen einem asymptotischen Grenzwert zustreben läßt, der zum Zeitpunkt der Beobachtung annähernd erreicht ist.

Die erste Annahme wird man häufig in der experimentellen Psychologie treffen müssen, wenn nämlich sofort nach der Manipulation einer unabhängigen Variablen die abhängige Variable erhoben wird. Treten die Effekte jedoch zeitverzögert auf, so würde das geprüfte experimentelle Kausalmodell fälschlich falsifiziert werden. Sinnvoll ist die erste Annahme wohl nur als zutreffend zu unterstellen, wenn ausdrücklich die sofortige Gedächtnisleistung, die sofort eintretenden Stimmungsveränderungen usw. analysiert werden sollen, wenn also der Validitätsanspruch für endogene Variablen von vornherein ausdrücklich auf den Zeitpunkt kurz nach der Manipulation der unabhängigen Variablen eingengt wird. Anderenfalls muß die Bedingung (2) oder (3) erfüllt sein. Allerdings läßt sich deren Zutreffen wohl selten überprüfen. In diesem Fall ist es ratsam, zu dynamischen Modellen überzugehen, die den Zeitparameter berücksichtigen und die endogenen Variablen zu mehreren Zeitpunkten enthalten.

Die Konzeption dynamischer Modelle führt insofern zu Problemen bei der Schätzung der Pfadkoeffizienten, als die impliziten Faktoren, welche die endogenen Variablen beeinflussen, nicht mehr als unkorreliert angesehen werden können. In Abb. 7 ist ein Modell mit einer exogenen Variablen zum Zeitpunkt $t-1$ (die Zeit wird als diskrete Variable behandelt) und zwei endogenen Variablen aufgeführt, die zu verschiedenen Zeitpunkten erhoben wurden. Auf den ersten Blick unterscheidet sich dieses Modell von dem in Abb. 3a nicht. Es besteht jedoch ein Unterschied: Da die endogene Variable zu verschiedenen Zeitpunkten erhoben wurde, wird eine Korrelation (sog. Autokorrelation) zwischen den impliziten Faktoren U_{t-1} und U_t angenommen. Bei mehr als zwei Meßzeitpunkten können auch Autokorrelationen höherer Ordnung entstehen (z.B. zwischen U_{t-2} und U_t). Wenn man nicht davon ausgehen kann, daß die Autokorrelationen Null sind, lassen sich die Pfadkoeffizienten auch nicht mehr so, wie bisher dargestellt, schätzen. Auf Modelle, die Autokorrelation einbeziehen, gehen Möbus und Nagl, die auch andere Möglichkeiten der Verlaufsanalyse untersuchen, in diesem Band ausführlich ein (vgl. auch Schubö et al. in Band 4 dieser Enzyklopädie); Verteilungsfreie Analysen von Zeitreihen behandelt ausführlich Lienert (1978)).

In der experimentellen Psychologie werden oftmals Veränderungen erfaßt, z.B. bei der Analyse von Lernkurven. Meistens wird eine varianzanalytische Auswertung vorgenommen, und es muß u.a. unterstellt werden, daß die Korrelationen zwischen den

Werten der abhängigen Variablen zu verschiedenen Meßzeitpunkten gleich groß sind. Verletzungen dieser Annahmen führen dazu, daß der F-Test zu häufig eine richtige Nullhypothese zurückweist. Diesen Fehler versucht man dadurch zu vermeiden, daß die Anzahl der Freiheitsgrade entsprechend des Ausmaßes der Heterogenität der Korrelationen zwischen den Meßwerten reduziert wird (vgl. dazu Huyng, 1978). Eine andere Möglichkeit besteht darin, die abhängigen Variablen nach bestimmten Regeln zu transformieren und multivariate Hypothesentests durchzuführen, die alle Hypothesen unter weniger restriktiven Annahmen zu prüfen gestatten, welche bei der Erfüllung der Voraussetzungen mit einer univariaten Varianzanalyse ebenfalls prüfbar wären. Es lassen sich also etwa auch Interaktionen zwischen einer oder mehreren Behandlungsvariablen mit dem Zeitfaktor auf Signifikanz prüfen. Eine ausgezeichnete Darstellung des Vorgehens findet sich bei McCall und Appelbaum (1973).

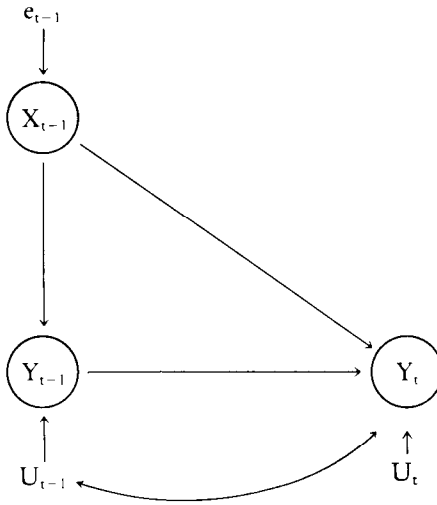


Abb. 7:

In dem Band der Enzyklopädie, dem dieses einführende Kapitel vorangestellt ist, findet sich schließlich eine Einführung in die Computer-Simulation psychischer Prozesse von Ueckert. Simulationsmodelle sind, wie der Autor betont, in der Regel als dynamische Modelle konzipiert. Ihre Besonderheit liegt darin, daß eine Theorie über den Gegenstandsbereich in eine Sprache übersetzt werden muß, die der Rechner versteht, so daß er „in allen auftretenden Situationen theoriegemäßes Verhalten zeigen kann. Dann kann das gleiche Experiment mit menschlichen und der „künstlichen“ Versuchsperson durchgeführt werden, und die Ergebnisprotokolle können miteinander verglichen werden“ (Deppe, 1977, 126).

Ueckert hebt die verschiedenen Datenquellen hervor, die für die Konzeption eines Simulationsmodells herangezogen werden können; dem Experiment komme wegen seiner methodologisch begründeten Künstlichkeit nur geringe Bedeutung zu. Hierin erblicken wir allerdings keine Besonderheit von Simulationsmodellen. Für die Generierung von experimentell zu prüfenden Hypothesen kommen auch alle möglichen Quellen in Frage, u.U. z.B. auch das einführende Verstehen (vgl. dazu Patzig, 1973). Die experimentelle Kontrolle wird erst bei der Prüfung der wie auch immer gefundenen Hypothese wichtig, um fälschliche Bestätigungen oder Widerlegungen zu vermeiden. Wie das obige Zitat von Deppe verdeutlicht, muß auch das Simulationsmodell überprüft werden, und wenn es sich auf menschliches Verhalten bezieht, so geschieht diese Prüfung durch den Vergleich des „Verhaltens“ des Computers mit dem Verhalten von Menschen. Deppe (1977) zeigt, wie Simulationsmodelle der Konzeptidentifikation aufgrund derartiger Vergleiche abgeändert wurden. Nach Ueckert sind Simulationsmodelle nicht falsifizierbar, da sie aus einer logischen und einer empirischen Komponente bestehen. Der logische Kern bleibt unabhängig davon, welche empirischen Daten auftreten, unbeeinflußt. Die empirische Komponente beinhaltet die Menge der intendierten Anwendungen eines Modells; deren Prüfung kann nur ergeben, ob ein Modell in einer bestimmten Situation anwendbar ist oder nicht. Dieser ursprünglich für hoch entwickelte physikalische Theorien formulierte sog. non-statement view von Sneed kann wohl nur mit Einschränkungen auf Theoriebildungen in der Psychologie übertragen werden (vgl. dazu Herrmann, 1976). Die Nicht-Anwendbarkeits-Interpretation erscheint für elaborierte und gut bewährte psychologische Theorien sinnvoll, wenn nicht eine andere Theorie vorhanden ist, deren Anwendbarkeitsbereich den des konkurrierenden Modells enthält, zusätzlich aber noch andere Anwendungen zuläßt. Liegt also der Fall vor, daß in diesem Sinne etwa die Theorie der Konzeptidentifikation von Levine (1975, Kap. 11 und 12) mit der von Bower und Trabasso (1964) verglichen werden kann, so wäre aufgrund der Falsifikationstheorie die letztgenannte als falsifiziert zu bezeichnen, während die von Levine als bewährt gelten kann. Durch die Formulierung beider Theorien in einer Sprache, die der Computer versteht, ändert sich hieran nichts.

Deppe (1977), der zu einer sehr ausgewogenen Beurteilung beim Vergleich von Simulationsmodellen mit mathematischen Modellen in der Psychologie gelangt, stellt zunächst heraus, daß beide eine psychologische Theorie voraussetzen. Sie sind kein Ersatz, sondern Hilfsmittel für die Theoriebildung. Wenn viele Größen zueinander in Beziehung stehen, können Simulationsmodelle die Komplexität häufiger als mathematische Modelle angemessen berücksichtigen; dadurch wird eine ganzheitliche Betrachtung wiederbelebt. Andererseits kann man aus einem mathematischen Modell exakte Prognosen für die *Population* der Menschen herleiten, die gemäß den Modellannahmen „funktionieren“. Dagegen sind Prognosen eines Simulationsmodells eine Stichprobe, die in un-

bekanntem Ausmaß verzerrt ist: „Wenn die Daten deutlich voneinander abweichen, bietet sich . . . die zusätzliche Möglichkeit, dies unter Hinweis auf den Stichprobencharakter der Prognosen zu entkräften“ (Deppe, 1977, 144). Mit anderen Worten: Mathematische Modelle sind strenger prüfbar als Simulationsmodelle, und dieser Gesichtspunkt, im Rahmen der Falsifikationstheorie betrachtet, ist sehr wichtig (vgl. dazu Hager und Westermann in diesem Band). Andererseits setzt dieses Kriterium voraus, daß die mathematische Modellbildung mit der Simulation verglichen werden kann, und wegen der besseren Handhabbarkeit komplexer Beziehungen durch Simulationsmodelle ist dieser Vergleich nicht immer möglich.

Literatur

- Bower, G. H. & Trabasso, T. 1964. Concept identification. In: Atkinson, R. C. (Ed.): Studies in mathematical psychology. Stanford: Stanford University Press.
- Brandstädter, J. 1976. Soziale Schicht, Umwelt und Intelligenz: Eine Pfadanalyse der Korrelationsbefunde von Marjoribanks. Psychologische Beiträge, 18, 35-53.
- Brandstädter, J. & Bernitzke, F. 1976. Zur Technik der Pfadanalyse. Ein Beitrag zum Problem der nichtexperimentellen Konstruktion von Kausalmodellen. Psychologische Beiträge, 18, 12-34.
- Bredenkamp, J. 1969. über die Anwendung von Signifikanztests bei theoriestendenden Experimenten. Psychologische Beiträge, 11, 275-285.
- Bredenkamp, J. 1972. Der Signifikanztest in der psychologischen Forschung. Frankfurt a. M.: Akademische Verlagsgesellschaft.
- Bredenkamp, J. 1980. Theorie und Planung psychologischer Experimente. Darmstadt: Steinkopff.
- Bredenkamp, J. 1982. Verfahren zur Ermittlung des Typs einer statistischen Interaktion. Psychologische Beiträge (im Druck).
- Campbell, D. T. & Stanley, J. C. 1973. Experimental and quasi-experimental designs for research on teaching. In: Gage, N. L. (Ed.): Handbook of research on teaching. Chicago: Rand McNally.
- Cohen, J. 1977². Statistical power analysis for the behavioral sciences. New York: Academic Press.
- Deppe, W. 1977. Formale Modelle in der Psychologie. Stuttgart: Kohlhammer.
- Gaensslen, H. & Schubö, W. 1973. Einfache und komplexe statistische Analyse. München: Reinhardt.
- Glaser, W. R. 1979. Statistische Entscheidungsprozeduren über Hypothesen in den Sozialwissenschaften. In: Albert, H. & Stapf, K. H. (Ed.): Theorie und Erfahrung. Stuttgart: Klett-Cotta.
- Herrmann, T. 1976. Die Psychologie und ihre Forschungsprogramme. Göttingen: Hogrefe.

- Hummell, H. J. & Ziegler, R. 1976. Zur Verwendung linearer Modelle bei der Kausalanalyse nicht-experimenteller Daten. In: Hummell, H. J. & Ziegler, R. (Ed.): Korrelation und Kausalität, Band 1, E5-E137. Stuttgart: Enke.
- Huyng, H. 1978. Some approximate tests for repeated measurement designs. *Psychometrika*, 43, 161-175.
- Kerlinger, F. N. & Pedhazur, E. J. 1973. Multiple regression in behavioral research. New York: Holt, Rinehart and Winston.
- Lehmann, G. 1980. Nichtlineare „Kausal-“ bzw. Dominanz-Analysen in Psychologischen Variablen-Systemen. *Zeitschrift für experimentelle und angewandte Psychologie*, 27, 257-276.
- Levine, M. 1975. A cognitive theory of learning. Research on hypothesis testing. Hillsdale: Erlbaum.
- Lienert, G. A. 1978. Verteilungsfreie Methoden in der Biostatistik. Band II. Meisenheim am Glan: Anton Hain.
- McCall, R. B. & Appelbaum, M. I. 1973. Bias in the analysis of repeated-measures designs: Some alternative approaches. *Child Development*, 44, 401-415.
- Patzig, G. 1973. Erklärung und Verstehen. Bemerkungen zum Verhältnis von Natur- und Geisteswissenschaften. *Neue Rundschau*, 3, 392-413.
- Popper, K. R. 1966². Logik der Forschung. Tübingen: Mohr.
- Rützel, E. 1979. Bayessches Hypothesentesten und warum die Bayesianer Bias-ianer heißen sollten. *Archiv für Psychologie*, 131, 211-232.
- Rützel, E. 1980. Korrektur zu E. Rützel: Bayessches Hypothesentesten und warum die Bayesianer Bias-ianer heißen sollten (Bayesianism or Biasianism?). *Archiv für Psychologie*, 132, 187-188.
- Wippich, W. & Bredenkamp, J. 1979. Bildhaftigkeit und Lernen. Darmstadt: Steinkopff.
- Witte, E. H. 1980. Signifikanztest und statistische Inferenz. Stuttgart: Enke.