

# Anwendung statistischer Modelle

Ingeborg Stelzl

Die Anwendung statistischer Methoden gehört für Psychologen zu den Selbstverständlichkeiten des Forschungsalltags. Einen Überblick über die Vielfalt statistischer Modelle zu geben, die für die Psychologie relevant sind, soll hier nicht versucht werden – das leistet das vorliegende Handbuch insgesamt. Ziel des folgenden Beitrags ist es vielmehr, anhand von typischen Beispielen die wesentlichen Schritte herauszuarbeiten und zu diskutieren, die in jeder Anwendung statistischer Modelle eine Rolle spielen, nämlich

- (1) die Formulierung des statistischen Modells,
- (2) die Herstellung des Bezugs zur inhaltlichen Fragestellung und zu den erhobenen Daten,
- (3) die Prüfung der Modellgeltung,
- (4) die Schätzung und eventuell Signifikanzprüfung der Parameter,
- (5) die Interpretation und die Suche nach rivalisierenden Erklärungen.

Diese Schritte können je nach Anwendungsfall in der Reihenfolge variieren und unterschiedlich stark in den Vordergrund treten. Die folgenden drei Beispiele unterscheiden sich durch eine unterschiedlich enge Verknüpfung zwischen statistischem Modell und inhaltlicher Theorie. Jedes davon soll eine Klasse von Anwendungsfällen repräsentieren:

Das erste Beispiel, ein *t*-Test, steht für einfache inferenzstatistische Verfahren. Statistisches Modell und inhaltliche Theorie sind relativ unabhängig. Dasselbe statistische Verfahren kann auf inhaltlich völlig unterschiedliche Fragestellungen angewendet werden. Als zweites Beispiel wird eine Pfadanalyse mit latenten Variablen behandelt: Durch das statistische Modell wird der inhaltlichen Theorienbildung ein allgemeiner Rahmen vorgegeben, der dann für das spezielle Problem adaptiert wird, wobei zugleich die inhaltliche Theorie und das formale Modell konkretisiert werden. Ähnliches gilt für andere lineare Strukturgleichungsmodelle wie z.B. die konfirmatorische Faktorenanalyse (vgl. Bentler, Wu & Houck; Rietz, Rudinger & Andres sowie Schönemann & Borg, in diesem Band), aber auch für Modelle der Testtheorie wie z.B. das *Latent-Class*-Modell (vgl. Langeheine & Rost, in diesem Band), linear-logistische Modelle (vgl. Roskam und Scheiblechner, in diesem Band) usw. Das zweite Beispiel steht für diese Gruppe von Modellen. Das letzte Beispiel stammt aus der Tradition der mathematischen Psychologie. Das mathematische Modell ist auf die spezielle

Fragestellung maßgeschneidert und ist Teil der Theorie selbst, da wesentliche Annahmen nur formal ausgedrückt und in ihrer inhaltlichen Bedeutung weitgehend offen gelassen sind.

## 1 Erstes Beispiel: der $t$ -Test

Das Interesse konzentriert sich hier auf Schritt (4), Parameterschätzung und Signifikanzprüfung. Alle anderen Schritte sind diesem Ziel untergeordnet und treten in den Hintergrund:

- Schritt (1), Formulieren des statistischen Modells, erübrigt sich, da das Modell schon fertig vorliegt.
- Schritt (2), Herstellen des Bezugs zur eigenen Fragestellung und den Daten, erfolgt bei der Festlegung des Auswertungsplans. Daß sich die Wahl des statistischen Verfahrens meist nicht zwingend aus den psychologischen Hypothesen und dem Versuchsplan ergibt, wurde von Westermann und Hager (1982) und Hager und Westermann (1983) und Hager (1992) betont. Sie weisen darauf hin, daß dieselben inhaltlichen Hypothesen in unterschiedlicher Weise in statistische Hypothesen umgesetzt werden können (multivariate oder univariate Auswertung, parametrische oder nicht-parametrische Verfahren usw.) und geben Empfehlungen für die Auswahl der Verfahren.
- Schritt (3), Prüfung der Modellgeltung, erfolgt, wenn die Voraussetzungen für die Anwendung des  $t$ -Tests überprüft werden. Wären z.B. die Beobachtungseinheiten einer Stichprobe nicht, wie im statistischen Modell vorausgesetzt, unabhängig gezogen, sondern z.B. den experimentellen Bedingungen ganze Schulklassen zugeordnet worden, so würde aus der Verwerfung der Nullhypothese (zwei unabhängige Stichproben aus derselben Normalverteilung) nicht die Alternativhypothese (zwei unabhängige Stichproben aus Normalverteilungen mit unterschiedlichen Erwartungswerten) folgen, sondern es würden weitere Alternativen (zwei Stichproben mit gleichen Erwartungswerten, aber Prüfgröße wegen Klumpeneffekten nicht  $t$ -verteilt) offen bleiben. Die Modellgeltung ist Voraussetzung jedes inferenzstatistischen Schlusses und muß deshalb vorher abgeprüft werden (Verteilungsvoraussetzungen) bzw. durch den Versuchsplan gewährleistet sein.
- Schritt (4), die Signifikanzprüfung für einen Parameter (Differenz der Erwartungswerte), steht beim  $t$ -Test eher im Zentrum des Interesses als die Prüfung der Modellgeltung. Näheres zur Durchführung dieser Signifikanzprüfung soll hier nicht ausgeführt werden. Eine ausführliche Darstellung findet man bei Willmes und bei Diepgen (in diesem Band).
- Schritt (5), die Interpretation, konzentriert sich dementsprechend darauf, das Zustandekommen der Mittelwertsdifferenz zu erklären (z.B. als experimentellen Effekt) und Alternativklärungen (z.B. durch Versuchsfehler) auszuschalten.

## 2 Zweites Beispiel: Pfadanalyse mit latenten Variablen

Das statistische Modell ist hier nicht schon fertig vorgegeben, sondern ist aus einem relativ weit gefaßten Modellrahmen anhand einer inhaltlichen Theorie erst noch zu entwickeln. Die ersten beiden Schritte, Formulierung des statistischen Modells und Anwendung auf ein inhaltliches Problem gehen dabei ineinander über.

Zur Veranschaulichung möge das in Abbildung 1 dargestellte hypothetische Modell dienen, das einen Spezialfall linearer Strukturmodelle darstellt:

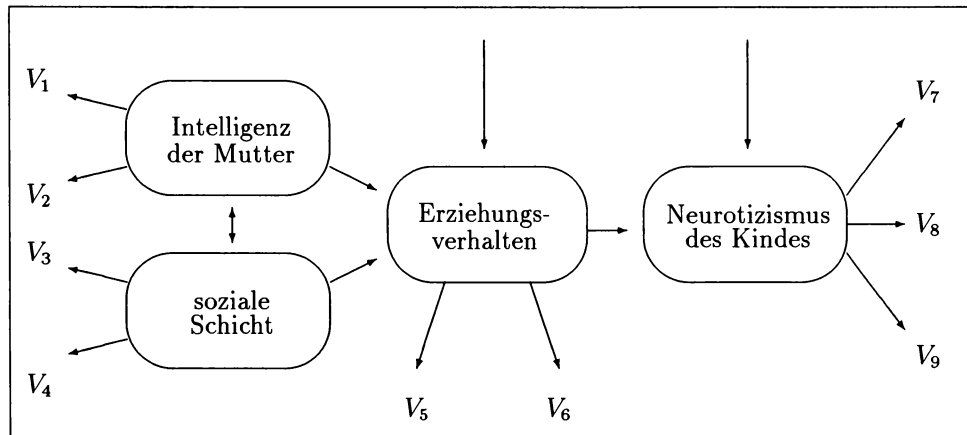


ABBILDUNG 1. Hypothetisches Pfadmodell aus vier latenten Variablen, die durch neun beobachtbare Variablen erfaßt werden.

Vier latente Variablen (Intelligenz der Mutter, Soziale Schicht, Erziehungsverhalten, Neurotizismus des Kindes) werden jede durch mehrere beobachtete Variablen ( $V_1$  bis  $V_9$ ) erfaßt, und dem Zusammenhang zwischen den latenten Variablen wird eine bestimmte Pfadstruktur unterstellt. Das Forschungsinteresse wird sich hier sowohl auf Schritt (3), die Prüfung der Gültigkeit des Modells (insbesondere der angenommenen Pfadstruktur) als auch Schritt (4), die Parameterschätzung (insbesondere auf die Pfadkoeffizienten als Schätzung für die Stärke des Einflusses) richten. Besondere Beachtung wird auch die Frage nach Alternativmodellen als rivalisierende Erklärungen für dieselben Daten (Schritt 5) finden müssen.

### 2.1 Prüfung der Modellgeltung

Man kann einmal prüfen, ob das Modell insgesamt mit den Daten vereinbar ist. Wenn man die entsprechenden Verteilungsannahmen als erfüllt voraussetzt, könnte man z.B. im vorliegenden Fall einen Likelihood-Quotienten-Test durchführen, der prüft, ob die aus dem Modell und den geschätzten Parametern rekonstruierte Kovarianzmatrix signifikant schlechter angepaßt ist als ein unbeschränktes Modell (alle Elemente der Kovarianzmatrix sind freie Parameter). Bei mangelnder Modellanpassung gibt ein solcher Globaltest allerdings keinen Hinweis darauf, welche Teile des Modells korrekturbedürftig sind.

Deshalb lohnt es sich zu überlegen, welche Teilstrukturen sich voneinander abgrenzen und getrennt prüfen lassen. Im vorliegenden Fall könnte man z.B. zunächst nur die Zuordnung der beobachteten Variablen zu den latenten durch eine konfirmatorische Faktorenanalyse (neun beobachtete Variablen, vier Faktoren, Null-Ladungen entsprechend den Hypothesen festgesetzt, freie Faktorenkorrelationen) testen.

Danach könnte man weiter fragen, ob das Pfadmodell mit den Korrelationen zwischen den Faktoren vereinbar ist (sechs Korrelationen werden durch vier Parameter des Pfadmodells angepaßt). Man kann aber auch die erste Hypothese (Zuordnung der beobachteten Variablen zu den latenten) noch weiter in Teilhypothesen aufspalten: Man kann prüfen, ob die Tests  $V_1$  und  $V_2$  – wie im Modell angenommen – kongenerisch sind, d.h. bis auf unterschiedliche Meßfehleranteile dasselbe messen (analog: die Tests 3 und 4, die Tests 5 und 6 sowie die Tests 7, 8, 9).

Ein komplexes Modell ist häufig nicht auf Anhieb an die Daten angepaßt, sondern bedarf mehrfacher Korrektur. Das birgt stets die Gefahr in sich, das Modell an bloß zufällige Eigenheiten der Daten anzupassen. Deshalb ist eine nochmalige Prüfung des korrigierten Modells an unabhängigen Daten (eine sog. „Kreuzvalidierung“) erforderlich – eine altbekannte, aber selten erfüllte Forderung.

## 2.2 Parameterschätzung und Signifikanzprüfung

Eine sinnvoll interpretierbare Parameterschätzung setzt voraus, daß

1. die Modellannahmen richtig sind. Daß das Modell mit den Daten vereinbar ist, ist dafür notwendig, aber nicht hinreichend (siehe Alternativmodelle).
2. das Modell vollständig ist. Wenn weitere, im Modell nicht enthaltene Variablen die zu erklärenden Zusammenhänge in den Daten mitbeeinflussen, kann es zu erheblichen Fehlschätzungen kommen.
3. die Lösung für die Parameter mathematisch eindeutig ist (Identifizierbarkeit der Parameter).

Wenn diese Voraussetzungen erfüllt sind, steht der Parameterschätzung und – bei bekannter Verteilung der Schätzwerte – einer Signifikanzprüfung eines Parameters (oder einer bestimmten Gruppe von Parametern) nichts im Wege. Vielfach ist jedoch nicht nur ein Parameter (bzw. eine fest vorgegebene Gruppe) zu testen, sondern eine ganze Reihe, wovon jeder Null oder von Null verschieden sein kann. Man hat es dann nicht mit einer Nullhypothese und einer Alternativhypothese zu tun, sondern mit einer ganzen Reihe von Modellvarianten, wobei die Schätzwerte für einen Parameter stark davon abhängen können, welche freien Parameter diese Modellvariante sonst noch enthält – ein aus der multiplen Regression und aus der nicht-orthogonalen Varianzanalyse wohlbekanntes Problem (s. oben: Die Kenntnis des richtigen Modells ist Voraussetzung einer sinnvoll interpretierbaren Parameterschätzung). Für lineare Strukturgleichungsmodelle geben Bentler und Bonett (1980) Entscheidungsstrategien an, um ein Modell mit möglichst wenig Parametern und möglichst guten Anpassungsindizes zu erhalten. Es handelt sich allerdings auch hier um post-hoc Strategien anhand der Daten, die eine Kreuzvalidierung erforderlich machen. Einen

Überblick über verschiedene Methoden zur Beurteilung der Modellgeltung, insbesondere verschiedene Indizes zur Güte der Modellanpassung, findet man bei Loehlin (1992).

### 2.3 Alternativmodelle als rivalisierende Erklärungen

Wenn man ein Modell entwickelt hat, das an die Daten gut angepaßt ist, so bedeutet das nicht, daß nicht an dieselben Daten andere Modelle ebensogut anpaßbar sein können. Es ist dann mit Hilfe der vorliegenden Daten keine Entscheidung zwischen den Modellen möglich. Das kann verschiedene Gründe haben:

(a) Die Modelle sind formal äquivalent, d.h. grundsätzlich empirisch ununterscheidbar. So sind z.B. die in Abbildung 2 gezeigten Pfaddiagramme äquivalent.

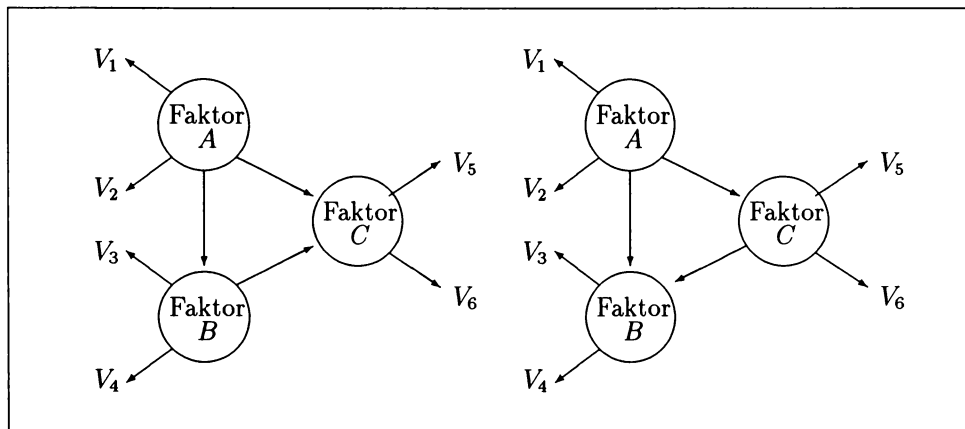


ABBILDUNG 2. Äquivalente Pfaddiagramme mit drei latenten Variablen und sechs beobachtbaren Variablen.

In beiden Modellen wird eine konfirmatorische Faktorenanalyse mit drei korrelierenden Faktoren angepaßt und die drei Korrelationen zwischen den Faktoren werden – in jeweils unterschiedlicher Weise – in drei Pfadkoeffizienten umgesetzt. Die beiden Modelle sind mit denselben Populations-Kovarianz-Matrizen vereinbar und werden für jede Stichproben-Kovarianz-Matrix dieselbe Anpassung liefern.

(b) Die Modelle sind nur bei Vorliegen spezieller Parameterkonstellationen ununterscheidbar. So kann es in einer konfirmatorischen Faktorenanalyse bei gegebenen Restriktionen im Ladungsmuster von der numerischen Konstellation der nicht fest vorgeschriebenen Ladungen abhängen, ob die Rotation fixiert ist oder nicht. Näheres dazu findet man bei Algina (1980) und Stelzl (1991). Ein Beispiel dazu ist in Tabelle 1 angegeben. Tabelle 1 zeigt ein hypothetisches Ladungsmuster für sechs beobachtete Variablen ( $V_1, V_2, \dots, V_6$ ) und drei Faktoren (I, II, III). Die Nullen sollen fest vorgeschrieben sein, der Buchstabe  $x$  steht für eine freie Ladung, die Faktorkorrelationen sollen freie Parameter sein. Die Ladungsmatrix ist nicht rotierbar, wenn die freien Ladungen die in Matrix  $A_I$  angenommenen Werte haben. Sie ist jedoch rotierbar, wenn z.B. Zeile 1 und 2 zueinander proportional sind, wie in Matrix  $A_{II}$ . Verwendet

man z.B. die angegebene Transformationsmatrix  $\mathbf{T}$ , so ergibt die Transformation der Ladungsmatrix zu  $\mathbf{A}_{II}^* = \mathbf{A}_{II} \mathbf{T}$  mit einer entsprechenden Transformation der Kovarianzmatrix der Faktoren  $\mathbf{M}_{II}$  zu  $\mathbf{M}_{II}^* = \mathbf{T}^{-1} \mathbf{M}_{II} (\mathbf{T}^{-1})'$  wieder eine Lösung, die alle im Ladungsmuster gesetzten Restriktionen erfüllt.

TABELLE 1. Nicht rotierbare und rotierbare Ladungsmatrix zum selben Ladungsmuster.

		Ladungsmuster		
		Faktoren		
		I	II	III
Variablen				
	$V_1$	0	x	x
	$V_2$	0	x	x
	$V_3$	x	0	x
	$V_4$	x	0	x
	$V_5$	x	x	0
	$V_6$	x	x	0

		nicht rotierbare	rotierbare	Transformationsmatrix	
		Ladungsmatrix $\mathbf{A}_I$	Ladungsmatrix $\mathbf{A}_{II}$	$\mathbf{T}$ zu $\mathbf{A}_{II}$	
		Faktoren	Faktoren		
$V_1$	0	.8	.5	$V_1$ 0 .8 .4	1 0 0
$V_2$	0	.6	.2	$V_2$ 0 .6 .3	-.5 1 0
$V_3$	.7	0	.1	$V_3$ .7 0 .3	1 0 1
$V_4$	.7	0	.3	$V_4$ .7 0 .3	
$V_5$	.5	.4	0	$V_5$ .5 .4 0	
$V_6$	.4	.5	0	$V_6$ .4 .5 0	

(c) Die Modelle können praktisch ununterscheidbar sein, weil ihre Vorhersagen so eng beisammen liegen, daß sie mit Daten, wie sie nach Art und Menge realistischweise erhoben werden können, nicht unterschieden werden können.

Für den an inhaltlichen Fragestellungen interessierten Forscher ist Unentscheidbarkeit zwischen Modellen kein befriedigendes Ergebnis. Er wird daher überlegen, welche zusätzliche Information zur Entscheidung führen könnte. Eine Entscheidung kann möglich werden, wenn das Modell durch weitere Variablen in geeigneter Weise ausgebaut wird. So z.B. sind die in Abbildung 2 dargestellten Modelle nicht mehr äquivalent, wenn eine weitere Variable  $D$  gefunden wird, von der bekannt ist, daß sie  $A$  und  $C$ , nicht aber  $B$  beeinflusst. Es entstehen die beiden in Abbildung 3 dargestellten Modelle, von denen das eine die Restriktion  $\rho(D, B | A) = 0$ , das zweite die

Restriktion  $\rho(D, B | A, C) = 0$  setzt, was nicht äquivalent ist. ( $\rho(X, Y | Z)$  bezeichnet hier die Korrelation der Zufallsvariablen  $X$  und  $Y$  für festgehaltene Ausprägungen von  $Z$ .) Näheres zur Äquivalenz von Pfadmodellen siehe Stelzl (1986), Breckler (1990), Lee und Hershberger (1990), MacCallum, Wegener, Uchino und Fabrigar (1993).

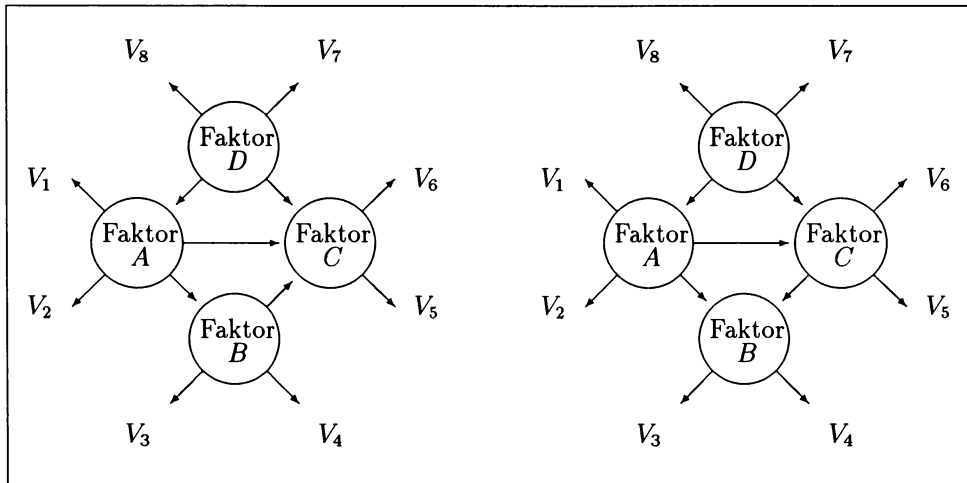


ABBILDUNG 3. Erweiterung der Pfadmodelle aus Abbildung 2 zu zwei nicht äquivalenten Modellen.

Im zweiten zur Uneindeutigkeit angeführten Beispiel, der konfirmatorischen Faktorenanalyse, bei der aufgrund der numerischen Parameterkonstellation noch Rotation möglich ist, läßt sich die Lösung eindeutig machen, wenn eine weitere beobachtete Variable mit einbezogen wird, die auf Faktor I ebenfalls eine vorgeschriebene Null-Ladung hat, aber zu den beobachteten Variablen  $V_1$  und  $V_2$  nicht kongenerisch ist, also z.B. auf Faktor III deutlich höher lädt als auf Faktor II. (Kongenerische Variablen haben bis auf unterschiedliche Meßfehleranteile gleiche faktorielle Zusammensetzungen, d.h. die ihnen entsprechenden Zeilen der Ladungsmatrix sind zueinander proportional).

Die diskutierten Beispiele für Alternativmodelle bewegten sich jeweils im selben Modellrahmen (Pfadanalyse, konfirmatorische Faktorenanalyse) wie das Ausgangsmodell. Das schließt natürlich nicht aus, daß andere Modellansätze zu praktisch ununterscheidbaren Vorhersagen führen. Beispiele dafür findet man in nahezu allen Bereichen der Psychologie.

### 3 Drittes Beispiel: ein Modell des Entscheidungsverhaltens

Als letztes soll ein Beispiel aus der Tradition der mathematischen Psychologie diskutiert werden, nämlich das von Aschenbrenner, Albert und Schmalhofer (1984) vorgestellte Modell für Wahlentscheidung. Das mathematische Modell ist hier auf das spezielle inhaltliche Problem maßgeschneidert. Wesentliche Annahmen der Theorie

betreffen unbeobachtbare kognitive Prozesse, die durch das formale Modell beschrieben werden, deren Realisierung (erlebnismäßig und physiologisch) aber offen bleibt.

Da das formale Modell nur im Zusammenhang mit dem zugehörigen Experiment verständlich wird, soll zunächst das Experiment kurz und vereinfacht dargestellt werden.

### *3.1 Das Experiment von Aschenbrenner et al. (1984)*

Den Vpn wurden Paare von Zeitschriften dargeboten, die auf einer Reihe von Dimensionen (z.B. Seitenzahl, Preis, Zahl der politischen Artikel usw.) beschrieben waren, und die Vp sollte sich jeweils für eine von beiden entscheiden. Insgesamt wurden aus neun Zeitschriften alle Paare gebildet und jedes Paar 20 mal dargeboten. Registriert wurde jeweils die getroffene Entscheidung und die hierfür benötigte Zeit. Außerdem wurden die Vpn gebeten, Ratings über die Wichtigkeit der Dimensionen abzugeben und – ebenfalls auf Rating-Skalen – die Ausprägungen auf den Dimensionen (z.B. Seitenzahl 300) zu bewerten.

### *3.2 Das Modell von Aschenbrenner et al. (1984)*

Aschenbrenner et al. nehmen folgenden Prozeß der Entscheidungsbildung an: Die Vp beachtet nicht alle Dimensionen, sondern wählt sie in der Reihenfolge ihrer subjektiven Wichtigkeit aus. Sie vergleicht dann die beiden Ausprägungen der zwei Zeitschriften und bildet die Differenz der entsprechenden Bewertungen. Die Differenzen werden so lange aufaddiert, bis ein Kriteriumswert überschritten wird. Bei Überschreiten des Kriteriumswerts wird die Entscheidung getroffen.

Aus diesem Grundmodell wurden verschiedene Varianten konstruiert: Es wurde ein probabilistisches Element eingeführt, indem entweder für die Bewertungen Zufallsschwankungen angenommen wurden, oder indem angenommen wurde, daß die Auswahl der als nächstes beachteten Dimension probabilistisch erfolgt. Darüber hinaus wurde probeweise angenommen, daß die Vp nicht Bewertungsdifferenzen aufaddiert, sondern einfach nur die Zahl der „Siege“ für jede Alternative vergleicht.

Aus den Modellannahmen, den Ratings über die Wichtigkeit der Dimensionen und den Bewertungen für die einzelnen Ausprägungen wurde dann vorhergesagt

- mit welcher Wahrscheinlichkeit sich die Vp jeweils für die eine oder andere Alternative entscheiden würde;
- wieviele Dimensionen die Vp bis zur Entscheidung verarbeiten müßte und daraus die relative Dauer der Entscheidungszeiten.

### *3.3 Prüfung der Modellgeltung und Ausschluß von Alternativmodellen*

Die Aussagen, die das Modell über die unbeobachtbaren kognitiven Prozesse macht, sind formal präzise und bei näherem Hinsehen psychologisch recht restriktiv. Das Modell impliziert u.a.,



- daß die Wichtigkeit einer Dimension und die Bewertung der Ausprägung (z.B. viele Farbbilder) nicht von anderen Dimensionen (z.B. Inhalt der Artikel) abhängt,
- daß das Abbruchkriterium gleich bleibt und nicht von der Zahl der bereits verarbeiteten oder noch nicht genutzten Dimensionen abhängt,
- daß eine Differenzregel verwendet wird und nicht z.B. das Verhältnis der Summen gebildet wird.

Demgegenüber erscheint die empirische Prüfung zunächst als recht schwach. Bei vorliegenden Ratings über die Wichtigkeit der Dimensionen und bekannten Bewertungen für die Ausprägungen dann die Entscheidungen vorherzusagen ist wohl von unterschiedlichen Modellansätzen her möglich. Auch die zweite Vorhersage, daß Entscheidungen zwischen auf vielen Dimensionen ähnlich bewerteten Alternativen länger dauern als zwischen unähnlichen, ist auf unterschiedliche Art begründbar und nicht nur aus dem vorliegenden Modell ableitbar. Diese Voraussagen allein, als globale Modelltests verstanden, erscheinen damit als unzulänglich, um die sehr restriktiven Modellannahmen zu belegen, und es stellt sich die Frage nach weiteren Prüfmöglichkeiten. Dazu bieten sich zwei Wege an: zum einen der Vergleich mit Alternativmodellen an denselben Daten, zum anderen der Versuch einer direkten Prüfung durch Variation der experimentellen Anordnung.

### Vergleich mit Alternativmodellen

Als *Alternativmodell* kann man einmal ein Modell heranziehen, das sich vom vorliegenden möglichst stark unterscheidet. Aschenbrenner et al. (1984) verglichen ihren Ansatz mit einem Modell, das annimmt, daß alle Dimensionen verarbeitet werden und die Bewertungen mit der Wichtigkeit gewichtet aufaddiert werden. Ein solcher Vergleich möglichst verschiedener Modelle sagt aber nicht, welche Unterschiede zwischen den Modellen für das bessere Abschneiden des einen oder anderen Modells verantwortlich sind. Von daher liegt es nahe, nur einzelne Annahmen auszutauschen. Das geschah bei Aschenbrenner et al. (1984), indem verschiedene Varianten des Grundmodells bezüglich ihrer Anpassung verglichen wurden. Solche Vergleiche setzen jedoch voraus, daß die allen Modellen gemeinsamen Annahmen richtig sind und haben insofern nur bedingte Aussagekraft. Der Vergleich mit Alternativmodellen ist also nur begrenzt geeignet, um einzelne Modellannahmen zu testen.

### Überprüfung einzelner Modellannahmen an zusätzlichen Daten

Im vorliegenden Beispiel erscheint z.B. die Annahme über die Unabhängigkeit der Bewertungen der einzelnen Dimensionen als einer direkten Überprüfung in einem gesonderten Experiment zugänglich. Die Anzahl und Reihenfolge der verarbeiteten Dimensionen wurde von Schmalhofer, Albert, Aschenbrenner und Gertzen (1986) und Böckenholt, Albert, Aschenbrenner und Schmalhofer (1991) in einer modifizierten Versuchsanordnung beobachtbar gemacht: Die Beschreibungen auf den Dimensionen wurden nicht mehr simultan dargeboten, sondern mußten von der Vp einzeln

abgefragt werden. Wenn auf diese Art ein Teil der Modellannahmen anderweitig abgestützt werden kann, sind bessere Voraussetzungen gegeben, die übrigen durch den Vergleich der globalen Anpassung rivalisierender Modelle zu testen und so schließlich die Modellgeltung zu belegen.

Das dargestellte Modell des Entscheidungsverhaltens dürfte in folgender Hinsicht für viele Modelle der mathematischen Psychologie typisch sein: Schritt (1) und (2), Formulieren des mathematischen Modells und Herstellen des Bezugs zur Fragestellung, sind ununterscheidbar. Das Hauptinteresse gilt der Prüfung der Modellgeltung (Schritt 3) und der Diskussion von Alternativmodellen (Schritt 5). Die Schätzung der Parameter (Schritt 4) ist nur Mittel zum Zweck.

#### 4 Fazit

Es wurden drei Beispiele diskutiert, von denen jedes einen bestimmten Einsatzbereich statistischer Modelle repräsentiert: Statistische Modelle können Hilfsmittel bei der Überprüfung von Theorien sein (*t*-Test), oder sie können einen bestimmten Rahmen setzen, innerhalb dessen eine inhaltliche Theorie formuliert wird (Pfadanalyse mit latenten Variablen). Schließlich können sie auch ein Teil der Theorie selbst sein (Modell für Wahlentscheidungen). In jedem der drei Beispiele waren alle fünf der eingangs genannten Schritte, die zur Anwendung statistischer Modelle erforderlich sind, enthalten, wenn auch mit unterschiedlichem Gewicht: Beim *t*-Test, der als Repräsentant klassischer Inferenzstatistik gewählt wurde, standen Parameterschätzung und Signifikanzprüfung im Vordergrund. Beim Beispiel 2, der Pfadanalyse mit latenten Variablen, traten auch die anderen Schritte deutlich in Erscheinung: Zunächst muß das statistische Modell vom allgemeinen Modellrahmen ausgehend anhand der inhaltlichen Fragestellung konkretisiert werden (Schritt 1 und 2: Formulieren des statistischen Modells und Herstellen des inhaltlichen Bezugs zu den Daten). Das Hauptinteresse gilt dann der Prüfung der Modellgeltung und dem Ausschluß von rivalisierenden Erklärungen, insbesondere auch von Alternativmodellen innerhalb desselben Modellansatzes (Schritt 3 und 5). Den Schätzwerten für die einzelnen Parameter (Schritt 4) kann – je nach Anwendungsbereich – größeres oder geringeres Gewicht zukommen. Im letzten Beispiel, dem Modell für Wahlentscheidungen, treten Modellentwicklung und Prüfung der Modellgeltung noch stärker in den Vordergrund, Parameterschätzung und Signifikanzprüfung von Parametern noch stärker zurück. Bei diesem Beispiel wurde auf Schritt 5, den Ausschluß rivalisierender Erklärungen, besonderes Gewicht gelegt, wobei neben der Anpassung von Alternativmodellen an dieselben Daten auch eine unabhängige experimentelle Prüfung einzelner Annahmen an neuen Daten erfolgte.

An den Beispielen sollte deutlich geworden sein, daß sinnvolle Parameterschätzungen und Signifikanzprüfungen ein gültiges statistisches Modell voraussetzen und inhaltliche Interpretationen den Ausschluß von Alternativerklärungen erfordern.

#### 5 Weiterführende Literatur

Das Buch von Bredenkamp (1972) führt in die Grundgedanken und die Logik der Signifikanzprüfung ein und diskutiert Anwendungsfälle unter wissenschaftstheoreti-

schen Gesichtspunkten. Als Ergänzung hierzu eignet sich Hagers (1992) Aufsatz, der das Verhältnis von psychologischer und statistischer Hypothese vertiefend behandelt, insbesondere die Ableitungs- bzw. Spezifikationschritte, die erforderlich sind, um von der Ebene der psychologischen Hypothese auf die der statistischen Hypothese zu gelangen und umgekehrt nach einer Entscheidung über die statistische Hypothese zu einer Entscheidung über die psychologische Hypothese zu kommen.

Das Buch von Loehlin (1992) kann als einfache, anwendungsorientierte Einführung in lineare Strukturgleichungsmodelle empfohlen werden.

## Literaturverzeichnis

- Algina, J. (1980). A note on identification in the oblique and orthogonal factor analysis models. *Psychometrika*, 45, 393–396.
- Aschenbrenner, K. M., Albert, D. & Schmalhofer, F. (1984). Stochastic choice heuristics. *Acta Psychologica*, 56, 153–166.
- Bentler, P. M. & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, 88, 588–606.
- Böckenholt, U., Albert, D., Aschenbrenner, K. M. & Schmalhofer, F. (1991). The effects of attractiveness, dominance, and attribute differences on information acquisition in multiattribute binary choice. *Organizational Behavior and Human Decision Processes*, 49, 258–281.
- Breckler, S. J. (1990). Applications of covariance structure modeling in psychology: Cause for concern? *Psychological Bulletin*, 107, 260–273.
- Bredenkamp, J. (1972). *Der Signifikanztest in der psychologischen Forschung*. Darmstadt: Steinkopf.
- Hager, W. (1992). Eine Strategie zur Entscheidung über psychologische Hypothesen. *Psychologische Rundschau*, 43, 18–29.
- Hager, W. & Westermann, R. (1983). Zur Wahl und Prüfung statistischer Hypothesen in psychologischen Experimenten. *Zeitschrift für experimentelle und angewandte Psychologie*, 30, 67–94.
- Lee, S. & Hershberger, S. (1990). A simple rule for generating equivalent models in covariance structure modeling. *Multivariate Behavioral Research*, 25, 313–334.
- Loehlin, J. C. (1992). *Latent variable models. An introduction to factor, path, and structural analysis*. 2. Aufl. Hillsdale: Erlbaum.
- MacCallum, R. C., Wegener, D. T., Uchino, B. N. & Fabrigar, L. R. (1993). The problem of equivalent models in applications of covariance structure analyses. *Psychological Bulletin*, 114, 185–199.
- Schmalhofer, F., Albert, D., Aschenbrenner, K. M. & Gertzen, H. (1986). Process traces of binary choices: Evidence for selective and adaptive decision heuristics. *The Quarterly Journal of Experimental Psychology*, 38A, 59–76.
- Stelzl, I. (1986). Changing a causal hypothesis without changing the fit: Some rules for generating equivalent path models. *Multivariate Behavioral Research*, 21, S. 309–331.
- Stelzl, I. (1991). Rival hypotheses in linear structure modeling: Factor rotation in confirmatory factor analysis and latent path analysis. *Multivariate Behavioral Research*, 26, 199–225.
- Westermann, R. & Hager, W. (1982). Entscheidungen über statistische und wissenschaftliche Hypothesen: Zur Differenzierung und Systematisierung der Beziehungen. *Zeitschrift für Sozialpsychologie*, 13, 13–21.