Von der Faktorenanalyse zu den Strukturgleichungsmodellen

Peter H. Schönemann und Ingwer Borg

Bei keiner der heute gebräuchlichen Methoden zur Datenanalyse läßt sich so deutlich aufzeigen wie bei der Faktorenanalyse und den Strukturgleichungsmodellen, wo ihre Ursprünge liegen: Die Faktorenanalyse, aus der sich später die Strukturgleichungsmodelle entwickelten, entstand als Modell zur Erklärung einer schon zu Beginn des Jahrhunderts bekannten empirischen Gesetzmäßigkeit im Bereich der Intelligenzforschung, dem Phänomen nämlich, daß die Leistungen von Personen bei verschiedenen Intelligenztest-Aufgaben im allgemeinen positiv korreliert sind: Wer die Aufgabe A gut löst, löst meist auch die Aufgabe B recht gut, und umgekehrt. Diese Regularität – von Guttman (1992) bezeichnet als „eines der best-bestätigten Phänomene in der gesamten Wissenschaft" und im sogenannten ersten Intelligenzgesetz präzisiert – verlangte nach einer Erklärung, möglichst natürlich nach einer, die sich empirisch überprüfen und gegebenenfalls auch widerlegen (falsifizieren) läßt.

Zu den ersten Erklärungsversuchen gehörte Spearmans Zwei-Faktoren-Theorie. Sie ist intuitiv einleuchtend und hatte wohl auch wegen der verwendeten Mathematik eine gewisse Aura der Wissenschaftlichkeit. Bei genauerer Betrachtung erwies sie sich aber als problematisch und nicht zwingend. Verschiedene Erweiterungen und Verallgemeinerungen des Grundgedankens, insbesondere die multiple Faktorenanalyse, wurden vorgeschlagen. Sie beseitigten allerdings die grundlegenden Probleme nicht, sondern führten im Gegenteil zu weiteren Fragen.

Gleichzeitig begann die Faktorenanalyse, sich von ihrer ursprünglichen Fragestellung, die sie nicht befriedigend beantworten konnte, abzulösen. Sie wurde zu einer Routine-Methode für die Analyse von Interkorrelationen, ohne sich allerdings vom formalen Ballast der ursprünglichen Fragestellung zu befreien. Die neueren Strukturgleichungsmodelle – zunächst, wie gehabt, euphorisch gefeiert (diesmal als Verfahren zur *Kausalmodellierung*) – tragen die Probleme der Faktorenanalyse in verschärfter Form fort.

1 Begriffe und Definitionen

Zunächst werden einige Begriffe und Definitionen eingeführt, die den Hintergrund der folgenden Darstellungen bilden (siehe auch Borg & Lingoes, 1987, Kapitel 17–19; Schönemann, 1981; Schönemann & Borg, 1983, Abschnitt 1.2; Searle, 1982; vgl. auch den Beitrag über Grundbegriffe der multivariaten Datenanalyse von Andres, in diesem Band):

https://doi.org/10.25521/HQM19

Sind y_i und y_j Zufallsvariablen, dann sind auch

$$u := ay_i \quad \text{und} \quad v := y_i + y_j, \quad a \in \mathbb{R} \tag{1}$$

Zufallsvariablen. Die Menge aller *Linearkombinationen* aus p Zufallsvariablen y_i,

$$V := \{v \mid v = \mathbf{a}'\mathbf{y} = \sum_i a_i y_i\}, \quad i = 1, \ldots, p; \quad a_i \in \mathbb{R}, \tag{2}$$
$$\text{wobei } \mathbf{y}' := (y_1, \ldots, y_p) \text{ und } \mathbf{a}' := (a_1, \ldots, a_p) \in \mathbb{R}^p,$$

ist somit ein Vektorraum unter den Verknüpfungen (1). V hat im allgemeinen die Dimension p. Er wird anschaulicher, wenn man ein Längen- und ein Winkelmaß einführt, so daß man sich jede Zufallsvariable als gerichtete Strecke („Pfeil") vorstellen kann. Dazu eignet sich als *Skalarprodukt*

$$s : V^2 \to \mathbb{R} : s(y_i, y_j) := E(y_i \cdot y_j), \tag{3}$$

wobei E den Erwartungswert bezeichnet. Haben alle y_i den Erwartungswert Null, so ist dieses Skalarprodukt die Kovarianz der Zufallsvariablen y_i und y_j, $Cov(y_i, y_j)$. Für $i = j$ ist $Var(y_i) := Cov(y_i, y_i)$. Geometrisch kann man die Standardabweichung als (euklidische) Länge des der Variablen y_i zugeordneten Pfeils deuten und die Korrelation, $Cor(y_i, y_j) := Cov(y_i, y_j)/[Var(y_i)Var(y_j)]^{1/2}$, als Cosinus des Winkels, den die Zufallsvariablen y_i und y_j einschließen.

Die $p \times p$ Varianz-Kovarianz-Matrix von p Zufallsvariablen $\mathbf{y}' := (y_1, \ldots, y_p)$ bezeichnet man mit $Var(\mathbf{y})$. Unterteilt man einen Vektor $\mathbf{y}$ von p Zufallsvariablen in zwei Untervektoren von m Zufallsvariablen u_i und n Zufallsvariablen v_j,

$$\mathbf{y}' := (\mathbf{u}'; \mathbf{v}') = (u_1, \ldots, u_m; v_1, \ldots, v_n), \tag{4}$$

dann enthält die $m \times n$ Untermatrix der $p \times p$ Matrix $Var(\mathbf{y})$

$$Cov(\mathbf{u}, \mathbf{v}) := (Cov(u_i, v_j)) \tag{5}$$

die Kovarianzen zwischen den ersten m Zufallsvariablen u_i mit den letzten $n = p - m$ Variablen v_j des Vektors $\mathbf{y}' = (y_1, \ldots, y_p)$.

Für die Kovarianzmatrix von zwei Mengen von Linearkombinationen, die mittels zweier Matrizen – einer $r \times m$ Matrix $\mathbf{A}$ und einer $s \times n$ Matrix $\mathbf{B}$ – aus den m Variablen u_i und den n Variablen v_j definiert sind, findet man

$$Cov(\mathbf{Au}, \mathbf{Bv}) = \mathbf{A}Cov(\mathbf{u}, \mathbf{v})\mathbf{B}'. \tag{6}$$

Ein Spezialfall ist

$$Var(\mathbf{Au}) = (Cov(\mathbf{a}_i'\mathbf{u}, \mathbf{a}_j'\mathbf{u})) = \mathbf{A}Var(\mathbf{u})\mathbf{A}'. \tag{7}$$

2 Spearmans Zwei-Faktoren-Theorie

Den Beginn der Faktorenanalyse kann man in Spearmans *Zwei-Faktoren-Theorie* sehen, die, wie man dem Titel seiner epochalen Arbeit „*General Intelligence, Objectively Measured and Determined*" (Spearman, 1904) entnehmen kann, der objektiven

Definition des verschwommenen und vielfach mißbrauchten Begriffs der Intelligenz dienen sollte. Spearman wollte die oben beschriebene Gesetzmäßigkeit der positiven Interkorreliertheit verschiedener Aufgaben von Intelligenztests erklären. Der intuitive Kern seiner Theorie ist, daß die Testleistungen bei jeder der p Aufgaben von jeweils zwei *Faktoren* bestimmt werden. Das heißt, daß der beobachtete Wert y_i in Aufgabe i erklärt wird als Linearkombination

$$y_i = a_i x + u_i z_i \qquad (i = 1, \dots p; \quad a_i > 0; \quad u_i \geq 0), \tag{8}$$

worin x ein *Faktor der allgemeinen Intelligenz* – von Spearman mit „g", für *general ability factor*, bezeichnet – ist, der mit unterschiedlichem, aber stets positivem Gewicht a_i zur Lösung aller Aufgaben beiträgt, und z_i ein zweiter, für den jeweiligen beobachteten Test y_i *spezifischer Faktor* ist, der mit dem Gewicht u_i nur in y_i und sonst keine andere Testaufgabe y_j eingeht. Alle p spezifischen Faktoren z_i sollen miteinander und mit x unkorreliert sein. Die stets positiven Beiträge von x erklären dann die positive Interkorreliertheit der Aufgaben untereinander. Die Aufgaben korrelieren andererseits nicht perfekt untereinander, weil zu ihrer Lösung im allgemeinen jeweils noch ein aufgabenspezifischer Faktor beiträgt.

Diese Modellannahmen kann man zusammenfassend formulieren als

$$\begin{aligned} \mathbf{y} &= \mathbf{a}x + \mathbf{U}\mathbf{z}, \\ Cov(x, \mathbf{z}) = 0, \quad Var(x) = 1, \quad Var(\mathbf{z}) &= \mathbf{I}_p, \\ \mathbf{U} = \text{ diagonal, mit } \quad u_{ii} &\geq 0. \end{aligned} \tag{9}$$

Standardisiert man die beobachteten Variablen y_i (so daß $Var(y_i) = 1$, $\mathbf{R} = Var(\mathbf{y})$), dann ist

$$1 = Var(y_i) = a_i^2 + u_i^2 = h_i^2 + u_i^2. \tag{10}$$

Das quadrierte Gewicht $h_i^2 := a_i^2$ des gemeinsamen Faktors nennt man die *Kommunalität* der i-ten Ausgangsvariablen. Sie zeigt den relativen Anteil des gemeinsamen Faktors in der i-ten Variablen an. Weiter folgt aus (9) und (10):

$$\begin{aligned} \mathbf{R} = Var(\mathbf{y}) &= \mathbf{a}\mathbf{a}' + \mathbf{U}^2, \\ \text{wobei } \mathbf{U}^2 = \text{diagonal}, \quad Cov(\mathbf{y}, x) &= \mathbf{a}. \end{aligned} \tag{11}$$

Die Gleichung $Cov(\mathbf{y}, x) = \mathbf{a}$ besagt, daß die Gewichte a_i die Kovarianzen (im standardisierten Fall also die Korrelationen) zwischen Ausgangsvariablen und dem gemeinsamen Faktor x sind (*Faktorenstruktur* von $\mathbf{y}$). Da in diesem Fall zudem $Var(x) = 1$ ist, sind die Elemente von $\mathbf{a}$ zugleich die Regressionsgewichte der Variablen y_i bezüglich des gemeinsamen Faktors x (*Faktorenmuster*). Wichtig ist, daß Spearmans Faktorenmodell unbedingt (d.h. ohne weitere Zusatzannahmen) falsifizierbar ist, sofern $p \geq 3$. Denn aus (9) und (10) folgt

$$\begin{aligned} Var(\mathbf{y}|x) &:= Var(\mathbf{y}) - Cov(\mathbf{y}, x) Var^{-1}(x) Cov(x, \mathbf{y}) \\ &= \mathbf{U}^2 = \text{diagonal}. \end{aligned} \tag{12}$$

Nach Auspartialisierung von x muß also die Kovarianzmatrix der Testaufgaben diagonal sein. Diese Bedingung ist nur dann trivial (d.h. für jedwede Beobachtung erfüllt), wenn $p \leq 2$. Anders ausgedrückt: Bei mehr als zwei Testaufgaben wird Spearmans Modell empirisch testbar.

Als Spearman diese Bedingung anhand zahlreicher, bereits veröffentlichter Arbeiten über Intelligenztests empirisch überprüfte, fand er sie durchweg bestätigt: „The average intercolumnar correlation from tables of 14 different investigators, summarizing 30 years of psychological researches and representing a great wealth of test material, was unity, as predicted by [Spearman's] unifocal hypothesis of a general factor. It seemed to be the most striking quantitative fact in the history of psychology" (Dodd, 1928, S. 214).

Dieser vermeintliche Durchbruch wurde auch von namhaften Vertretern des IQ-Geschäfts wohlwollend vermerkt: „As to the existence of 'g' as a common factor, there seems to be no possibility of doubt. Psychometrics, without it, loses its basic prop" (Wechsler, 1939, S. 8).

Dieser Schein trog jedoch. Einmal abgesehen davon, daß sich der Grad der Bestätigung der Spearmanschen Vorhersagen im allgemeinen als doch nicht so gut erwies und zudem die Elimination („*explaining away*"; Guttman, 1992, S. 182) verschiedener Testaufgaben erforderte, fand Thomson (1916), daß man genau dieselben Vorhersagen über die Form der beobachteten Kovarianzmatrix erhält, wenn man, statt von nur *einem*, von *unendlich vielen* Intelligenzfaktoren ausgeht (*sampling theory of intelligence*). Statt (9) sei folgendes Modell angenommen:

$$y_i = \sum_{k \in K \subset J} a_k x_k, \qquad \{J = 1, \ldots, m\}, \tag{13}$$
$$\text{also } \mathbf{y} = \mathbf{A}\mathbf{x}, \quad Var(\mathbf{x}) = \mathbf{I}_m, \quad m \gg p,$$

wobei die x_k eine Zufallsstichprobe von m_i Faktoren darstellen, die aus einer sehr großen Anzahl untereinander unkorrelierter *Intelligenzfaktoren* $x_1, \ldots, x_m$ gezogen wurde; $\mathbf{A}$ ist eine $p \times m$ Matrix, die viel mehr Spalten als Zeilen hat, und die in jeder Zeile m_i Einsen und $m - m_i$ Nullen in Zufallsverteilung enthält. Aus (13) folgt, daß

$$Var(\mathbf{y}) = \mathbf{A}\mathbf{A}'. \tag{14}$$

Wenn $m \to \infty$, dann wird die Korrelation zwischen der i-ten und j-ten Intelligenztest-Aufgabe

$$\begin{aligned} r_{ij} &= Cov(y_i, y_j)/[Var(y_i)Var(y_j)]^{1/2} \\ &= \mathbf{a}_i'\mathbf{a}_j/[(\mathbf{a}_i'\mathbf{a}_i\mathbf{a}_j'\mathbf{a}_j)]^{1/2} \\ &= p_i p_j/(p_i p_j)^{1/2}, \end{aligned} \tag{15}$$

wobei $p_i = m_i/m$. Mit der Definition $\mathbf{p}' := (p_1^{1/2}, \ldots, p_p^{1/2})$ wird offenkundig, daß die Korrelationsmatrix $\mathbf{R} = (r_{ij})$ der p y_i,

$$\mathbf{R} = \mathbf{p}\mathbf{p}' + \mathbf{I} - \text{Diagonale von } \mathbf{p}\mathbf{p}', \tag{16}$$

mit (11) identisch ist. Also ist Spearmans Faktorenmodell, das nur einen Intelligenzfaktor postuliert, von dem von Thomson, das unendlich viele solcher Faktoren postuliert, empirisch nicht zu unterscheiden. Ein Zahlenbeispiel hierfür findet man bei Schönemann (1981, S. 364).

Wilson (1928) bemerkte zudem in einer Rezension der „*Abilities of Man*“ (Spearman, 1927) gleichsam am Rande, daß in Spearmans Faktorenmodell ein Fehlschluß steckt, den Spearman, und nicht nur er, völlig übersehen hatte: Die dieses Modell definierende Grundgleichung (9) *erklärt* p beobachtete Variablen durch $p+1$ miteinander unkorrelierte *latente Faktoren*. Zwar ist nur einer davon, x $(=g)$, von vorrangigem Interesse: Er soll zur „objektiven Bestimmung der allgemeinen Intelligenz“ dienen. Genau das aber wird unmöglich gemacht durch die Zusatzbedingung, daß alle $p+1$ Faktoren unkorreliert sein sollen, und erst sie macht Spearmans Modell falsifizierbar (im Unterschied zur oberflächlich verwandten, aber empirisch nicht testbaren Hauptkomponentenanalyse; siehe z.B. Schönemann & Steiger, 1976; vgl. auch den Beitrag über Grundlagen der multivariaten Datenanalyse von Andres, in diesem Band). Es entsteht damit das Problem, wie man aus nur p Testvektoren $p+1$ linear unabhängige Vektoren – die Faktoren – bestimmen soll. Das Problem ist mathematisch unlösbar, auch wenn das nicht allgemein offensichtlich zu sein scheint (siehe Mallows & Tukey, 1982, S. 149).

Piaggio (1933) veröffentlichte eine Formel, mit der man die vom Faktorenmodell postulierten unkorrelierten Faktoren x_1, x_2, ..., x_{p+1} aus den p beobachteten Variablen und einer *zusätzlichen*, mit ihnen unkorrelierten, sonst aber völlig aus der Luft gegriffenen Variablen s perfekt konstruieren kann:

$$x = \mathbf{a}'\mathbf{R}^{-1}\mathbf{y} + ks, \; \mathbf{z} = \mathbf{U}\mathbf{R}^{-1}\mathbf{y} - k\mathbf{U}^{-1}\mathbf{a}s, \tag{17}$$
$$\text{wobei } k^2 := Var(x|\mathbf{y}) = 1 - \mathbf{a}'\mathbf{R}^{-1}\mathbf{a}, \quad Var(s) = 1, \quad Cov(s, \mathbf{y}) = \mathbf{0}.$$

Als Maß der Unbestimmbarkeit eines Faktors schlug Guttman (1955) die kleinste Korrelation, $r_{min} := min[Cor(x, x^*)] = 2\mathbf{a}'\mathbf{R}^{-1}\mathbf{a} - 1$, zwischen zwei äquivalenten Lösungen, x und x^*, vor. Ist etwa die multiple Korrelation zwischen den Ausgangsvariablen y_i und dem gemeinsamen Faktor x gleich .707, dann ist $r_{min}(x, x^*) = 0$. (Für weitere numerische Beispiele siehe Schönemann, 1983, 1987; für empirische Werte für r_{min} siehe Schönemann & Wang, 1972, S. 78-85.)

Dieses Problem der *Faktorenunbestimmbarkeit* wurde in den 30er Jahren in England und in den USA von namhaften Mathematikern und Statistikern (u.a. Heywood, Wishart, Camp, Irwin, siehe Steiger & Schönemann, 1978) ausgiebig diskutiert, denn ihnen war klar, daß es für den von Spearman angemeldeten Anspruch, den Begriff der *Intelligenz* objektiv bestimmt zu haben, nämlich als den allen Intelligenztest-Aufgaben gemeinsamen Faktor x $(=g)$, von ausschlaggebender Bedeutung ist.

3 Spezialfall: Klassische Testtheorie

Ein besonders einfacher, aber wichtiger Spezialfall ist die sog. Klassische Testtheorie, die man aus dem Spearmanschen Modell (9) erhält, wenn man $a_i = a$ setzt:

$$\mathbf{a}' = (a, a, \ldots, a), \quad \text{so daß auch} \quad \mathbf{U}^2 = u^2\mathbf{I}_p. \tag{18}$$

Die Variable $t := ax$ nennt man *wahren Wert* (*true score*), die Variablen $e_1 := uz_1$ und $e_2 := uz_2$ *Fehler*.

$$y_1 = t + e_1, \quad y_2 = t + e_2, \quad Cov(t, e_k) = Cov(e_1, e_2) = 0 \tag{19}$$

sind dann zwei *perfekt parallele Tests*. Ihre Korrelation

$$r_{yy} := Cov(y_1, y_2)/Var(y_1) = Var(t)/Var(y_1) \tag{20}$$

nennt man *Reliabilität* (Zuverlässigkeit; vgl. Stumpf, in diesem Band).

Da die *true-score*-Variable t und die beiden Fehlervariablen e_i ebenfalls nicht eindeutig definiert sind, kann man fragen, wieso sie zur *Erklärung* der tatsächlich beobachteten Tests beitragen können. Solange man diese *Theorie*, wie zumeist üblich, auf nur zwei beobachtbare (*perfekt parallele*) Tests beschränkt, ist sie zudem auch nicht falsifizierbar. In diesem Zusammenhang mag interessieren, daß die gegenwärtig recht populäre *Item-Response-* (vormals: *Latent-Trait-*) Theorie (vgl. Roskam und Scheiblechner, in diesem Band) ebenfalls an einem Uneindeutigkeitsproblem krankt (Cressie & Holland, 1983, S. 139), das bisher anscheinend übersehen wurde (z.B. Lord, 1980; Fischer, 1974).

4 Verallgemeinerungen

4.1 Multiple Faktorenanalyse: $m \geq 2$

Man erhält eine triviale Verallgemeinerung, wenn man, statt nur einen, m (≥ 2) gemeinsame Faktoren postuliert:

$$\mathbf{y} = \mathbf{Ax} + \mathbf{Uz}, \quad \mathbf{A} \in \mathbb{R}^{p \times m}, \quad Cov(\mathbf{x}, \mathbf{z}) = \mathbf{0}, \quad Var(\mathbf{z}) = \mathbf{I}_p. \tag{21}$$

Dieser simple Gedanke wird oft Thurstone zugeschrieben (z.B. von Thurstone, 1947, S. 333), obwohl er sich bereits bei Garnett (1919) findet. Die Formel (10) führt nun zu

$$Var(\mathbf{y}) = \mathbf{APA}' + \mathbf{U}^2, \quad Cov(\mathbf{y}, \mathbf{x}) = \mathbf{AP}, \tag{22}$$

wobei $\mathbf{P} := Var(\mathbf{x})$ die Varianz-Kovarianz-Matrix der m gemeinsamen Faktoren $x_1, \ldots, x_m$ bezeichnet, die man gewöhnlich standardisiert (so daß die Diagonale von $\mathbf{P}$ aus Einsen besteht).

Als erste geringfügige Komplikation muß man nun die Faktorenstruktur (22) vom Faktorenmuster

$$Cov(\mathbf{y}, \mathbf{x}) Var^{-1}(\mathbf{x}) = \mathbf{A} \tag{23}$$

unterscheiden. Eine zweite Komplikation ergibt sich aus

$$\begin{aligned} \mathbf{y} &= \mathbf{Ax} + \mathbf{Uz} = (\mathbf{AT})(\mathbf{T}^{-1}\mathbf{x}) + \mathbf{Uz} = \mathbf{A}^*\mathbf{x}^* + \mathbf{Uz}, \\ & \forall\, \mathbf{T} \in \mathbb{R}^{m \times m} \quad \text{mit Rang}(\mathbf{T}) = m. \end{aligned} \tag{24}$$

Ist also $(\mathbf{A}, \mathbf{U}, \mathbf{P}, \mathbf{x}, \mathbf{z})$ eine Lösung für einen beobachteten Vektor $\mathbf{y}$ von p Beobachtungsvariablen, dann ist auch $(\mathbf{A}^*, \mathbf{U}, \mathbf{P}^*, \mathbf{x}^*, \mathbf{z})$ eine Lösung mit $\mathbf{x}^* := \mathbf{Tx}$, $\mathbf{A}^* := \mathbf{AT}^{-1}, \mathbf{P}^* := \mathbf{TPT}^{-1}$ für alle regulären $m \times m$ Matrizen $\mathbf{T}$ (*Rotationsproblem*). Wenn $m > 1$ ist, dann gibt es somit unendlich viele gleich gute Lösungen für die gemeinsamen Faktoren $\mathbf{x}$.

Thurstone schlug vor, die Matrix $\mathbf{T}$ so zu wählen, daß jede der p Zeilen des Faktorenmusters $\mathbf{A}^*$ möglichst viele Nullen enthält (*Einfachstruktur*). Dieser Vorschlag war ursprünglich als empirische Hypothese gemeint, die von Fall zu Fall zu überprüfen ist (Thurstone, 1947, S. 363). Ein von Bargmann (1955) entwickelter statistischer Test für diese a priori keineswegs selbstverständliche Hypothese wurde jedoch, wie es scheint, von den meisten Faktorenanalytikern mit Bedacht gemieden, denn „es ist keine Übertreibung zu sagen, daß ungefähr die Hälfte der Forscher ihre Schlußfolgerungen annulliert fänden — was vermutlich der Grund ist, warum Signifikanztests der Einfachstruktur so selten veröffentlicht werden" (Cattell, 1978, S. 175). Mittlerweile wird das Einfachstruktur-Kriterium (formuliert als *Varimax*-Kriterium) in der faktorenanalytischen Praxis routinemäßig angewendet. Auf der Strecke bleibt dabei „die wichtige Frage, warum man Einfachstruktur in vielen Systemen von Variablen [überhaupt] erwarten sollte" (Steiger, 1994, S. 204). Im Grunde ist die Einfachstrukturhypothese wenig plausibel: Sie „spannt den Wagen vor das Pferd und behauptet, daß es unmöglich sei, bestimmte Testaufgaben zu konstruieren" (Guttman, 1992, S. 186), solche Testaufgaben nämlich, die Nicht-Null-Ladungen auf allen m Faktoren haben. Daß man „faktoriell reine" Testaufgaben gerne hätte oder konstruieren *könnte*, ist eine andere Frage.

In der ersten Version seines Faktorenanalysebuchs (Thurstone, 1935), das, wie man seinem Titel „*The Vectors of Mind*" entnehmen kann, noch weitgehend inhaltlichen Fragestellungen gewidmet war, schlug Thurstone noch ein weiteres Rotationskriterium vor: Die *positive manifold hypothesis*. Damit stellte er die Frage, ob sich ein $\mathbf{T}$ finden läßt, mit dem $\mathbf{x}$ so transformiert werden kann, daß $\mathbf{Tx} = \mathbf{x}^*$ nur noch nicht-negative Elemente hat. Wäre das möglich, ließen sich die nicht-negativen Interkorrelationen der Testaufgaben in Intelligenztests als Folge nicht-negativer Beiträge der gemeinsamen Faktoren erklären. Diese Frage, 1935 noch breit diskutiert in einem eigenen Kapitel, wurde 1947 in dem nun zu „*Multiple Factor Analysis*" umbenannten Buch nur noch kurz erwähnt und verschwand dann ganz aus der Literatur, weil sich die Faktorenanalyse von der ursprünglichen Fragestellung ganz ablöste und von Thurstone gar als „allgemeine wissenschaftliche Methode" empfohlen wurde.

Mit dieser Ablösung verschwanden allerdings nicht die formalen Probleme. So erweiterten Kestelman (1952) und Guttman (1955) die Konstruktion von Piaggio (17) auf den m-dimensionalen Fall:

$$\mathbf{x} = \mathbf{PA'R}^{-1}\mathbf{y} + \mathbf{Ks}, \quad \mathbf{z} = \mathbf{UR}^{-1}\mathbf{y} - \mathbf{U}^{-1}\mathbf{AKs}, \quad Var(\mathbf{x} \mid \mathbf{y}) = \mathbf{KK'}. \tag{25}$$

Definiert man als überschlägiges Maß für alle $p+m$ Faktoren das Mittel ihrer Minimalkorrelationen, dann findet man, daß dieses Mittel im unkorrelierten Fall ($\mathbf{P} = \mathbf{I}$, *orthogonale Rotation*) eine von den Daten völlig unabhängige Konstante ist (Schönemann, 1971),

$$r_{min} = (p-m)/(p+m), \tag{26}$$

die um so kleiner wird, je mehr Faktoren m extrahiert werden. Empirische Werte hierzu berichten Schönemann und Wang (1972, S. 75-77).

Heute weiß man, daß die von Thurstone eingeläutete Ära der „explorativen Faktorenanalyse" wenige Einsichten von Bestand gebracht hat. Die scheinbar geringfügige Erweiterung von einem auf m gemeinsame Faktoren, von Thurstone und seinen Jüngern als Durchbruch gepriesen, erwies sich im Nachhinein als Rückschritt, denn damit wurde das Falsifikationsproblem relativiert: Wenn ein Faktor die Daten nicht beschrieb, dann eben zwei und ganz sicher mehr als zwei.

4.2 Statistische Verfahren: Uneingeschränkte Maximum-Likelihood-Faktorenanalyse (UMLFA)

Es lag nahe zu versuchen, diesem Verfall wissenschaftlicher Standards mit statistischen Verfahren Einhalt zu gebieten. Der ihnen zugrundeliegende Gedanke knüpft direkt an Spearman an: Will man wissen, wieviele Faktoren zu extrahieren sind, dann kann man den Schätzwert der partiellen Korrelationsmatrix,

$$Cor(\mathbf{y}|\mathbf{x}) = \mathbf{U}^{-1}[Var(\mathbf{y}) - \mathbf{AA}']\mathbf{U}^{-1}, \tag{27}$$

einem statistischen Test unterwerfen, der prüft, ob sie mit der Nullhypothese einer Einheitsmatrix verträglich ist. Die hierzu benötigten Schätzverfahren bauen meist auf dem Maximum-Likelihood-Prinzip auf und sind iterativ (etwa Lawley, 1940).

Als Nebenprodukt erhält man einen (asymptotischen) *Likelihood-Ratio*-Test (siehe z.B. Haagen & Seifert, 1979, Kap. II), mit dem man die Hypothese $H_0 : m = m_0$ statistisch überprüfen könnte, wenn er nur einigermaßen trennscharf wäre. Computersimulationen zeigen aber, daß er für mäßige Stichprobengrößen die Hypothese, daß m klein ist, favorisiert, wobei die *Power* eines Tests zum Niveau $\alpha = 5\%$ unter .1 liegen kann (Schönemann, 1981). Das Teststärkeproblem wurde jedoch von den Statistikern ausgeklammert. So blieb der Anschein gewahrt, solche Tests ermöglichten eine objektive Bestimmung der Anzahl der gemeinsamen Faktoren.

4.3 Konfirmatorische Faktorenanalyse

Statt nur die Anzahl m der gemeinsamen Faktoren statistisch zu überprüfen, kann man ebensogut bestimmte Elemente der Kovarianzparameter des Faktorenmodells, z.B. eine Untermenge der $p \cdot m$ Parameter a_{ik}, per Nullhypothese festlegen und die übrigen *freien* Parameter dann iterativ schätzen (*restricted maximum likelihood factor analysis*; siehe etwa Anderson & Rubin, 1956). Mit diesem „konfirmatorischen“ Ansatz kann man z.B. das Rotationsproblem statistisch angehen. Eine gegenwärtig recht populäre Verallgemeinerung ist das sogenannte LISREL-Modell (vgl. auch Bentler, Wu & Houck sowie Rietz, Rudinger & Andres, in diesem Band).

4.4 Das LISREL-Modell

Beim LISREL-Modell wird das Faktorenmodell mit der von Wright (1920, siehe auch Li, 1975) konzipierten *Pfadanalyse* verknüpft. Die hiermit eröffnete Möglichkeit der *Kausalmodellierung* (Blalock, 1971) mit *latenten* (*unmeasured*) Variablen, den Faktoren nämlich, wurde schnell so populär, daß Guttman (1977) anmerkte: „[...] according to current journals [...] sociologists keep discovering new fundamental path frameworks every month; and sociological graduate students are required routinely to hand in, as individual class exercises, new discoveries equalling Gregor Mendel's“ (S. 104).

Kausalmodelle haben die Form

$$\mathbf{y} = \mathbf{Cy} + \mathbf{Ax} + \mathbf{Vz}, \quad Cov(\mathbf{x}, \mathbf{z}) = \mathbf{0}, \quad Var(\mathbf{z}) = \mathbf{I}_p. \tag{28}$$

Im einfachsten, *rekursiven* Fall hat man etwa

$$
\begin{aligned}
y_1 &= a_{11}x_1 + \ldots + a_{1m}x_m + e_1, \\
y_2 &= a_{21}x_2 + \ldots + a_{2m}x_m + e_2 + c_{21}y_1, \\
\ldots & \\
y_p &= a_{p1}x_1 + \ldots + a_{pm}x_m + e_p + c_{p1}y_1 + \ldots + c_{p,p-1}y_{p-1}; \\
\\
\mathbf{y} &= \mathbf{Ax} + \mathbf{Uz} + \mathbf{Cy}.
\end{aligned} \tag{29}
$$

Hierbei werden p abhängige (*endogene*) Variablen y_i als Linearkombinationen von m unabhängigen (*exogenen*) Variablen x_k und p Fehlervariablen $e_i := u_i z_i$ erklärt, wobei y_i auch von den $i-1$ vorausgehenden y_j $(j < i)$ abhängen darf. Die Definitionen $b_{ij} := -c_{ij}$ für $i \neq j$ und $b_{ii} := 1$ führen zu der Darstellung

$$\mathbf{By} = \mathbf{Ax} + \mathbf{Vz}, \quad Cov(\mathbf{x}, \mathbf{z}) = \mathbf{0}, \quad Var(\mathbf{z}) = \mathbf{I}_p, \tag{30}$$

wobei im allgemeineren, *nicht-rekursiven* Fall **B** und **V** nur regulär zu sein brauchen. Für rekursive Modelle ist die Matrix **B** eine untere Dreiecksmatrix mit Einsen in der Diagonalen und **V** = **U** = diagonal. Ohne Zusatzbeschränkungen hat dieser Spezialfall stets eine einfache exakte Lösung (Schönemann, 1984), ist also nicht falsifizierbar.

Im nicht-rekursiven Fall dürfen auch die Fehler in $\mathbf{e} := \mathbf{Vz}$ miteinander korrelieren, da **V** im Gegensatz zu **U** nicht länger auf diagonale Matrizen beschränkt ist. Der Wegfall dieser Beschränkung kann sich auf die Interpretation der substantiell wichtigeren Parameter **A** und **B** drastisch auswirken. Das illustriert van de Geer (1971) an einem schönen Beispiel: Zunächst postuliert er eine Kausalkette, die von der Erziehung des Vaters über die des Sohns und eine weitere Variable zum endgültigen Beruf des Sohns führt (Modell I, rekursives Pfadmodell). In diesem Fall ist das Regressionsgewicht der Endvariablen (Beruf des Sohns) auf die Erziehung des Vaters -1. Van de Geer (1971) kommentiert: „Ein naiver Soziologe könnte leicht versucht sein, dieses Regressiongewicht kompensatorisch zu deuten [...]. Söhne wollen anders sein als ihre Väter. Tatsächlich aber wurden die Korrelationen [die Modell I perfekt beschrieb, d. Verf.] von einem ganz anderen Modell generiert“ (S. 125), einem Modell II nämlich, das von korrelierten Fehlern ausgeht und bei dem das Regressionsgewicht der Endvariablen auf die Erziehung des Vaters keineswegs -1, sondern genau 0 war.

Wichtig ist, daß Wright (1920) die y_i und x_k als beobachtbar voraussetzte. Ersetzt man sie durch *latente Variablen* im Sinne des Faktorenmodells, dann nimmt man in Kauf, daß sie alle uneindeutig werden (Vittadini, 1989). Eben das schlugen Turner und Stevens (1959) vor, „(...) die Möglichkeit, die Faktorenanalyse mit (...) der Analyse von Regressiongewichten [d.h. Pfadanalyse, d. Verf.] zu verbinden“ (Nachdruck in Blalock, 1971, S. 88).

Jöreskogs (1973) LISREL-Programm verwirklicht diesen Vorschlag im Rahmen der Maximum-Likelihood-Methode. Dieses Modell wird definiert durch

$$
\begin{aligned}
&\mathbf{By} = \mathbf{Ax} + \mathbf{Uz}, \quad \mathbf{B} \text{ regulär}, \quad Cov(\mathbf{x}, \mathbf{z}) = \mathbf{0}, \quad Var(\mathbf{z}) = \mathbf{I}_n, \\
&\mathbf{y}^* = \mathbf{My} + \mathbf{e}, \quad \mathbf{x}^* = \mathbf{Nx} + \mathbf{d}, \quad Cov(\mathbf{y}, \mathbf{e}) = \mathbf{0}, \quad Cov(\mathbf{x}, \mathbf{d}) = \mathbf{0}.
\end{aligned} \tag{31}
$$

Die erste Zeile in (31) definiert ein *Strukturmodell* zwischen *n latenten Variablen* y_r und m latenten Variablen x_s im Sinne von Wrights Pfadanalyse. Die zweite Zeile definiert zwei *Meßmodelle*, die die beobachteten Variablen, $y_1^*, \ldots, y_p^*$ und $x_1^*, \ldots, x_q^*$ im Sinne des Faktorenmodells mit n latenten y_r und m latenten x_s *erklären* sollen.

Daß keine der *latenten Ursachen* eindeutig definiert ist, wird im LISREL-Manual taktvoll verschwiegen. Tatsächlich verschlimmert sich das Unbestimmbarkeitsproblem im LISREL-Fall, weil man dort meist wesentlich mehr *latente Variablen* postuliert als im Modell der multiplen Faktorenanalyse, wo m bewußt klein gehalten wurde. Zusätzlich hat man es nun auch noch mit einer erheblich größeren Zahl von Parametern zu tun $(\mathbf{A}, \mathbf{U}, \mathbf{B}, \mathbf{M}, \mathbf{N}, Var(\mathbf{x}), Var(\mathbf{y}), Var(\mathbf{e}), Var(\mathbf{d}))$, die alle mehr oder minder willkürlich festgesetzt oder iterativ geschätzt werden dürfen. Die praktisch schrankenlose Plastizität des LISREL-Modells – selbst im Vergleich zum Modell der multiplen Faktorenanalyse – unterhöhlt nicht nur den Anspruch auf statistische Inferenz, sie fordert fast zum Mißbrauch auf, denn bei hinreichender Geduld wird sich schon irgendein „Kausalmodell" finden, das für die gerade vorliegenden Daten statistisch nicht abgelehnt werden muß. Hat man ein solches Modell gefunden, das die Daten mit einer akzeptablen Anpassungsgüte erklärt, dann kann man dieses Modell zudem nach Regeln von Stelzl (1986) und Lee und Hershberger (1990) - z.B. durch die Umkehrung gewisser kausaler Pfade – in verschiedenste Modellvarianten transformieren, die alle den gleichen Fit haben. Außerdem existieren im allgemeinen noch weitere Modelle, die „praktisch" den gleichen Fit haben.

5 Weiterführende Literatur

Zur Faktorenanalyse gibt es zahlreiche Lehrbücher, in denen die Faktorenanalyse allerdings meist als unproblematisches Verfahren der Datenanalyse dargestellt wird. Oft fehlt eine klare Abgrenzung zur Hauptkomponentenmethode. Als moderner Klassiker zur Faktorenanalyse gilt Mulaik (1972). Inhaltliche Forschung im Rahmen der Faktorenanalyse wird recht ausführlich in Pawlik (1968) dargestellt. Zur Einführung in die Strukturgleichungsmodelle wird meist Bollen (1989) verwendet. Eine gute Übersicht und eine umfangreiche Literaturliste zu neueren Entwicklungen und Perspektiven der kausalen Modellierung findet sich in Faulbaum und Bentler (1994).

Literaturverzeichnis

Anderson, T. W. & Rubin, H. (1956). Statistical inference in factor analysis. In: *Proceedings of the 3rd Berkeley Symposium on Mathematical Statistics and Probability*, *5*, 111.

Bargmann, R. (1955). Signifikanzuntersuchungen der einfachen Struktur in der Faktorenanalyse. *Mitteilungsblatt für Mathematische Statistik*, *8*, 1–14.

Blalock, H. M. (1971). *Causal models in the social sciences.* New York: Aldine.

Bollen, K. A. (1989). *Structural equations with latent variables.* New York: Wiley.

Borg, I. & Lingoes, J. C. (1987). *Multidimensional similarity structure analysis.* New York: Springer.

Cattell, R. B. (1978). *The scientific use of factor analysis in behavioral and life sciences.* New York: Plenum.

Cressie, N. A. C. & Holland, P. W. (1983). Characterizing the manifest probabilities of latent trait models. *Psychometrika*, *48*, 129–141.

Dodd, S. C. (1928). The theory of factors, I., II. *Psychological Review*, *35*, 211–234, 261–279.

Faulbaum, F. & Bentler, P. M. (1994). Causal modeling: Some trends and perspectives. In I. Borg & P. P. Mohler (Eds.), *Trends and perspectives in empirical social research* (S. 224–249). New York: De Gruyter.

Fischer, G. H. (1974). *Einführung in die Theorie psychologischer Tests.* Bern: Huber.

Garnett, J. C. M. (1919) General ability, cleverness and purpose. *British Journal of Psychology*, *9*, 345–366.

van de Geer, J. P. (1971). *Introduction to multivariate analysis for the social sciences.* San Francisco: Freeman.

Guttman, L. (1955) The determinacy of factor score matrices with implications for five other basic problems of common-factor theory. *British Journal of Statistical Psychology*, *8*, 65–81.

Guttman, L. (1977). What is not what in statistics. *The Statistician*, *26*, 81–107. Wiederabgedruckt in I. Borg (Ed.) (1981). *Multidimensional data representations: When and why.* Ann Arbor: Mathesis Press. Ebenfalls wiederabgedruckt in S. Levy (Ed.) (1994), *Louis Guttman on theory and methodology: Selected writings.* Dartmouth: Aldershot.

Guttman, L. (1992). The irrelevance of factor analysis for the study of group differences. *Multivariate Behavioral Research*, *27*, 175–204.

Haagen, K. & Seifert, H. G. (1979). *Methoden der Statistik für Psychologen.* Stuttgart: Kohlhammer.

Jöreskog, K. G. (1973). A general method for estimating a linear structural equation system. In A. S. Goldberger & O. D. Duncan (Eds.) *Structural equation models in the social sciences* (pp. 85–107). New York: Seminar Press.

Kestelman, H. (1952). The fundamental equation of factor analysis. *British Journal of Psychology*, *5*, 1–6.

Lawley, D. N. (1940). The estimation of factor loadings by the method of maximum likelihood. *Proceedings of the Royal Society of Edinburgh*, *60*, 64–82.

Lee, S. & Hershberger, S. (1990). A simple rule for generating equivalent models in covariance structure modeling. *Multivariate Behavioral Research*, *25*, 313–334.

Li, C. C. (1975). *Path analysis: A primer.* Pacific Grove: Boxwood.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems.* Hillsdale: Erlbaum.

Mallows, C. L. & Tukey, J. W. (1982). An overview of techniques of data analysis, emphasizing its exploratory aspects. In T. Y. Oliviera & B. Epskin (Eds.) *Some recent advances in statistics* (pp. 111–172). London: Academic Press.

Mulaik, S. A. (1972). *The foundations of factor analysis.* New York: McGraw.

Pawlik, K. (1968). *Dimensionen des Verhaltens.* Stuttgart: Huber.

Piaggio, H. T. H. (1933). Three sets of conditions necessary for the existence of a *g* that is real and unique except in sign. *British Journal of Psychology*, *24*, 88–105.

Schönemann, P. H. (1971). The minimum average correlation between equivalent sets of uncorrelated factors. *Psychometrika*, *36*, 21–30. Erratum: *Psychometrika*, *38*, 149.

Schönemann, P. H. (1981). Factorial definitions of intelligence: Dubious legacy of dogma in data analysis (pp. 325–374). In I. Borg (Ed.) *Multidimensional data representations: When and why.* Ann Arbor: Mathesis.

Schönemann, P. H. (1983). Do IQ tests really measure intelligence? *Behavioral and Brain Sciences*, *6*, 311–315.

Schönemann, P. H. (1984). An algebraic solution for the recursive path model with manifest variables. *Bulletin of the Psychonomic Society*, *22*, 455–458.

Schönemann, P. H. (1987). Jensen's *g*: Outmoded theories and unconquered frontiers. In S. Modgil & C. Modgil (Eds.)*Arthur Jensen: Consensus and controversy* (pp. 313–327). Barcombe Lewes: Falmer Press.

Schönemann, P. H. & Borg, I. (1983). Grundlagen der mehrdimensionalen Skaliermethoden. In H. Feger & J. Bredenkamp (Hrsg.) *Messen und Testen.* (= Enzyklopädie der Psychologie, Thembereich B, Serie I, Band 3, S. 257–344). Göttingen: Hogrefe.

Schönemann, P. H. & Steiger, J. H. (1976). Regression component analysis. *British Journal of Mathematical and Statistical Psychology, 29*, 175–189.

Schönemann, P. H. & Wang, M. M. (1972). Some new results on factor indeterminacy. *Psychometrika, 37*, 61–91.

Searle, S. R. (1982). *Matrix algebra useful for statistics.* New York: Wiley.

Spearman, C. (1904). General intelligence, objectively determined and measured. *American Journal of Psychology, 15*, 201–293.

Spearman, C. (1927). *The abilities of man.* New York: MacMillan.

Steiger, J. H. (1994). Factor analysis in the 1980's and the 1990's: Some old debates and some new developments. In I. Borg & P. P. Mohler (Eds.) *Trends and perspectives in empirical social research* (pp. 201–223). New York: De Gruyter.

Steiger, J. H. & Schönemann, P. H. (1978). A history of factor indeterminacy. In S. Shye (Ed.) *Theory construction and data analysis in the behavioral sciences* (pp. 136–178). San Francisco: Jossey Bass.

Stelzl, I. (1986). Changing a causal hypothesis without changing the fit: Some rules for generating equivalent path models. *Multivariate Behavioral Research, 23*, 297–326.

Thomson, G. (1916). A hierarchy without a general factor. *British Journal of Psychology, 8*, 271–281.

Thurstone, L. L. (1935). *The vectors of mind.* Chicago: University of Chicago Press.

Thurstone, L. L. (1947). *Multiple factor analysis.* Chicago: University of Chicago Press.

Turner, M. E. & Stevens, C. D. (1959). The regression analysis of causal paths. *Biometrics, 15*, 236–258. Wiederabgedruckt in H. M. Blalock (1971). *Causal models in the social sciences.* New York: Aldine.

Vittadini, G. (1989). Indeterminacy problems in the LISREL model. *Multivariate Behavorial Research, 24*, 397–414.

Wechsler, D. (1939). *The measurement of adult intelligence.* Baltimore: Williams and Wilkins.

Wilson, E. B. (1928). Review of „The Abilities of Man, their Nature and Measurement" by C. Spearman. *Science, 67*, 244–248.

Wright, S. (1920). The relative importance of heredity and environment in determining the Piebald pattern of guineapigs. *Proceedings of the National Academy of Sciences, 6*, 320–332.