

Hildebrandt, Lutz; Klapper, Daniel

**Working Paper**

## Möglichkeiten und Ansätze der Analyse dreimodaler Daten für die Marktforschung

SFB 373 Discussion Paper, No. 1997,90

**Provided in Cooperation with:**

Collaborative Research Center 373: Quantification and Simulation of Economic Processes, Humboldt University Berlin

*Suggested Citation:* Hildebrandt, Lutz; Klapper, Daniel (1997) : Möglichkeiten und Ansätze der Analyse dreimodaler Daten für die Marktforschung, SFB 373 Discussion Paper, No. 1997,90, Humboldt University of Berlin, Interdisciplinary Research Project 373: Quantification and Simulation of Economic Processes, Berlin,  
<https://nbn-resolving.de/urn:nbn:de:kobv:11-10064674>

This Version is available at:

<https://hdl.handle.net/10419/66281>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

# Möglichkeiten und Ansätze der Analyse dreimodaler Daten für die Marktforschung

Lutz Hildebrandt und Daniel Klapper  
Institut für Marketing II  
Wirtschaftswissenschaftliche Fakultät  
Humboldt-Universität zu Berlin  
10178 Berlin

24. Oktober 1997

## 1 Einleitung

Die in der Marktforschungspraxis durchgeführten Datenanalysen beruhen im allgemeinen auf der stufenweisen Anwendung verschiedener Techniken zur Verdichtung der in den Daten enthaltenen Information und der Suche nach Einflußstrukturen<sup>1</sup>. Üblich sind z. B. häufig die Kombination von informationsverdichtenden Verfahren wie Faktorenanalyse oder mehrdimensionale Skalierung mit der Clusteranalyse sowie Verfahren zur Analyse von Beziehungen wie Regressionsanalyse oder Conjoint Analyse mit der Clusteranalyse zur Aufdeckung von latenten Strukturen und den darin vorhandenen Abhängigkeiten bzw. Einflußstrukturen. Die simultane Analyse von Daten mit Verfahrenskombinationen wie sie beispielsweise in der Strukturgleichungs-Methodologie (LISREL, EQS, AMOS) in der Verbindung von Faktorenanalyse und Regressionsanalyse erfolgt, ist bisher eher die Ausnahme.

Die in diesem Beitrag vorgestellte Methodik geht von der Dreidimensionalität (Stimuli, Skalen, Personen) des üblicherweise analysierten Datenkörpers aus. Jede dieser Dimensionen erfaßt eine Einflußgröße (Modalität), die in mehrere latente Faktoren, die für die Ausprägung der Beobachtungsdaten ursächlich sind, zerlegt werden kann. Grundprinzip ist die simultane Faktorenanalyse über die drei Modalitäten, wodurch die Datenverdichtung, die Aufdeckung von Strukturen und die Segmentierung miteinander verbunden werden.

---

<sup>1</sup>Diese Arbeit wurde durch den Sonderforschungsbereich 373 "Quantifikation und Simulation Ökonomischer Prozesse" an der Humboldt-Universität zu Berlin unterstützt.

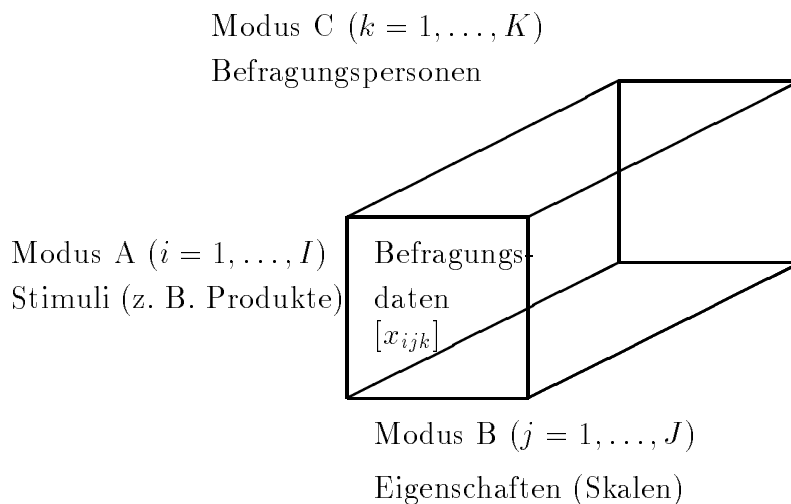
Die Grundidee dieses Verfahrens wurde schon in den 60er Jahren von Tucker (1963, 1966) entwickelt. Erst heute wird aber durch die Fortschritte der PC-Technik eine praktische Anwendung in der Marktforschung - mit geringen Rechenzeiten - möglich. Der vorliegende Beitrag führt in die Methodik der dreimodalen Analyse von Daten ein, erläutert deren Möglichkeiten und dokumentiert ein praktisches Anwendungsbeispiel aus der Marktforschung - die Segmentierung nach Wettbewerbsstrukturen im Spülmittelbereich. Es werden mehrere Modellansätze dargestellt und eine Anwendung des Programmpakets TUCK-ALS3 dokumentiert.

## 2 Dreimodale Daten in der Marktforschung

Eine der grundlegenden Aufgaben explorierender multivariater Analysemethoden in der Marktforschung ist die Verdichtung von Information und die Aufdeckung von Hintergrundvariablen oder Faktoren, die die Ausprägung der Daten erzeugt haben. Sind diese Variablen theoretisch begründbar, so spricht man von latenten Variablen oder theoretischen Konstrukten. Ein typisches Marketingbeispiel ist die Anwendung der Faktorenanalyse auf Einstellungsskalen über Stimuli wie Produktmarken, die in einer Stichprobe potentieller Käufer beurteilt wurden. Die Daten liegen dabei als dreidimensionaler Datenkörper vor, dessen Dimensionen durch drei Modalitäten gebildet werden, Skalen, Stimuli und Personen. Abbildung 1 zeigt die Struktur des dreidimensionalen und auch dreimodalen Datenkörpers. Ein Element  $x_{ijk}$  ist ein Eintrag der dreimodalen Datenmatrix mit  $i = 1, \dots, I$  Stimuli,  $j = 1, \dots, J$  Eigenschaften (Skalen) und  $k = 1, \dots, K$  Befragungspersonen. Die Matrix  $\mathbf{X}_k$  definiert darin eine  $(I \times J)$ -dimensionale Matrix, die die Elemente der  $k$ -ten Person(enscheibe) enthält. Die üblicherweise zur Analyse solcher Daten eingesetzten Verfahren sind die Faktorenanalyse, die Hauptkomponentenanalyse und die mehrdimensionale Skalierung (z. B. Backhaus, Plinke, Erichson & Weiber 1996). Deren Anwendung setzt jedoch i. d. R. Inputdaten mit einer zweidimensionalen und zweimodalen Struktur voraus. Man geht dabei im allgemeinen von einer Homogenität im Urteilsverhalten der Befragungspersonen aus. Bei den meisten Befragungen in der Marktforschung ist aber die Annahme homogener Urteile nicht gerechtfertigt.

Zur Verdichtung der Information des vorliegenden dreidimensionalen (und dreimodalen) Datenkörpers beschränkt man sich zudem üblicherweise nur auf eine Modalität des Datenkörpers (etwa bei Imageskalen, um die Basisdimensionen zu ermitteln, nach denen die Stimuli beurteilt wurden). So kann durch die Aggregation oder Nichtberücksichtigung des Modus C in der Faktorenanalyse in der allgemeinen Modellgleichung  $x_{ij} = \sum_{s=1}^S a_{is}b_{js} + e_{ij}$  der Index  $k$  für die dritte Dimension und Modalität entfallen. Häufig wird deshalb nach der Analyse einer Modalität das Ergebnis einer weiteren Analyse unterzogen, z. B. mit einer Clusteranalyse. Diese stufenweise Vorgehensweise setzt wiederum voraus, daß zwischen den einzelnen Modalitäten keine Interaktionen bestehen. Liegen Interaktionen vor, würden bei der Analyse dreimodaler Daten mit Verfahren, die nur die Informa-

Abbildung 1: Der dreidimensionale Datenarray



tionen von einer oder zwei Modalitäten berücksichtigen, wichtige Informationen verloren gehen, die für das Verständnis der Gesamtstruktur der Daten bedeutsam sind. Notwendig ist deshalb eine simultane Analyse aller Modalitäten, die die Interaktionen zwischen der latenten Struktur in den Personen (z. B. Segmente) oder latenten Wettbewerbsclustern mit den Imagekomponenten explizit berücksichtigen. Hier setzt der methodische Ansatz dieses Beitrags an.

Die Leistungsfähigkeit moderner Rechner erlaubt es heute, flexible Ansätze der Datenanalyse einzusetzen, die eine simultane Schätzung und systematische Repräsentation aller wichtigen Beziehungen in den dreimodalen Daten ermöglichen (vgl. z. B. Hildebrandt & Klapper). Das methodische Fundament dieser Analysetechniken beruht auf den Arbeiten von L.R. Tucker (1963, 1964, 1966, 1972), die die Grundlagen für die dreimodale Faktorenanalyse und die Integration individueller Unterschiede in die mehrdimensionale Skalierung bilden.

### 3 Die Analyse dreimodaler Daten

#### 3.1 Modellansätze der dreimodalen Datenanalyse

Ausgehend von unterschiedlichen Techniken in der Faktorenanalyse können bei der dreimodalen Analyse zwei Forschungsrichtungen unterschieden werden. Eine Forschungsrichtung basiert auf der direkten Erweiterung der Analyse von Kovarianzstrukturen, in denen die Rohdaten als stochastisch aufgefaßt werden. Hier sind Ansätze von Bloxom (1968), Bentler & Lee (1978), McDonald (1978), Browne (1984) und Bentler, Poon & Lee (1988) entwickelt worden. Die Kompliziertheit dieser Methoden und die Ungeklärtheit von Iden-

tifikationsproblemen haben jedoch bisher zu einer geringen Akzeptanz bei Forschungsaufgaben geführt, so daß bisher alle Anwendungen dieser Methoden auf die oben genannten Quellen beschränkt bleiben. Die zweite Forschungsrichtung ist praktikabler und basiert im wesentlichen auf einer Erweiterung der Hauptkomponentenmethode. Hier werden die Modalitäten als nicht-stochastisch betrachtet, sondern die Daten lassen sich als Grundgesamtheit interpretieren. Seit Tuckers Beitrag (1966) sind in dieser Forschungsrichtung eine Vielzahl von Methoden entwickelt worden, u. a. von Harshman (1970), Carroll & Chang (1970), Kroonenberg & de Leeuw (1980), Kiers, Kroonenberg & ten Berge (1992) sowie Lohmöller (1984).

Obwohl die Methoden auf Grundlage der Hauptkomponentenmethode in der dreimodalen Faktorenanalyse zu den eingeführten Verfahren in der Psychometrie gehören und das Verfahren in Gebieten der Chemometrie, Geographie, Linguistik, Pädagogik und Soziologie erfolgreich eingesetzt wurde, sind bisher Anwendungen auf Marktforschungsprobleme die Ausnahme. Der hier vorgestellte Ansatz baut auf der zweiten methodischen Richtung auf und erlaubt die Erweiterungen der Hauptkomponentenanalyse wie sie grundlegend zuerst von Tucker (1966) entwickelt und dokumentiert wurden.

### 3.2 Das Tucker3-Modell

Tucker war der erste, der den klassischen Ansatz der Faktorenanalyse (i. e. S. den der Hauptkomponentenanalyse) auf den Fall dreimodaler Daten erweitert hat. Sein Modell basiert auf der Singulärwertzerlegung (z. B. Eckart & Young 1936) und einer Generalisierung dieser Zerlegung für den dreimodalen Fall mit Hilfe des Kronecker-Produktes. Das dreimodale Hauptkomponentenmodell nach Tucker (1966) wird Tucker3-Modell genannt und hat die folgende Grundstruktur:

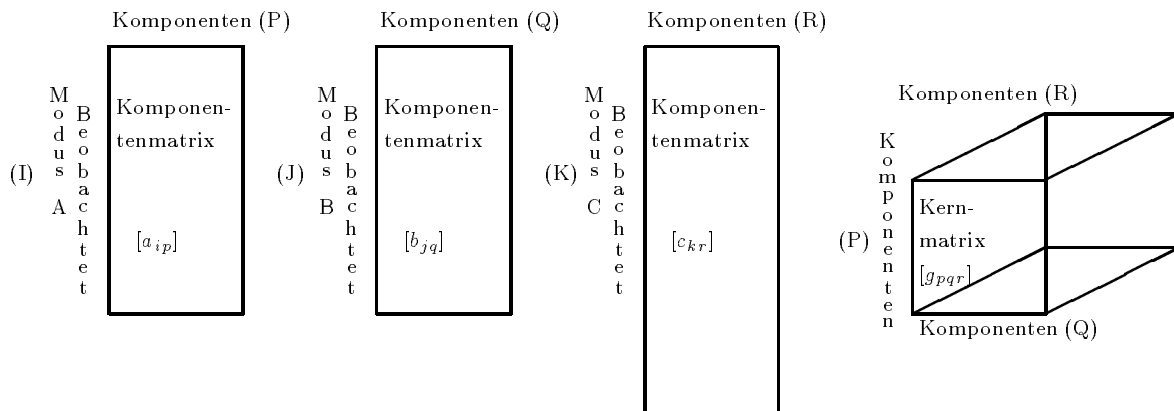
$$x_{ijk} = \hat{x}_{ijk} + e_{ijk} \quad (1)$$

$$\hat{x}_{ijk} = \sum_{p=1}^P \sum_{q=1}^Q \sum_{r=1}^R a_{ip} b_{jq} c_{kr} g_{pqr}. \quad (2)$$

Die Koeffizienten  $a_{ip}$ ,  $b_{jq}$  und  $c_{kr}$  sind Elemente der  $(I \times P)$ -,  $(J \times Q)$ - und  $(K \times R)$ -dimensionalen Komponentenmatrizen  $\mathbf{A}$ ,  $\mathbf{B}$  und  $\mathbf{C}$ .  $g_{pqr}$  ist ein Element der sogenannten Kernmatrix  $\mathbf{G}$ . Die Kernmatrix enthält die Informationen darüber, wie die einzelnen Komponenten der Matrizen  $\mathbf{A}$ ,  $\mathbf{B}$  und  $\mathbf{C}$  miteinander verknüpft sind. Die Dimension der Kernmatrix  $\mathbf{G}$  ergibt sich entsprechend über die Anzahl der Komponenten der Modi  $\mathbf{A}$ ,  $\mathbf{B}$  und  $\mathbf{C}$ .  $e_{ijk}$  gibt die Abweichung bei der Anpassung von  $x_{ijk}$  durch die drei Komponentenmatrizen und die Kernmatrix an.

Im allgemeinen wird die Datenmatrix nicht in die maximale Anzahl an Komponenten  $I$ ,  $J$  und  $K$ , sondern nur in  $P$ ,  $Q$  und  $R$  Komponenten mit  $P < I$ ,  $Q < J$  und  $R < K$  dekomponiert. Die Festlegung der Größen von  $P$ ,  $Q$  und  $R$  erfolgt unter dem Gesichtspunkt der Bedeutsamkeit der Komponenten und der erklärten Varianz in den Daten. Die

Abbildung 2: Die Struktur des Tucker3-Modells



Struktur des Tucker3-Modells ist in Abbildung 2 wiedergegeben.

Die Matrix **A** enthält die  $P$  Komponenten, welche die Produktbeurteilungen der Marken zusammenfassen. Ähnliche Komponentenladungen zweier Marken in einer Komponente der Matrix **A** belegen, daß die Marken vergleichbar beurteilt werden. Entsprechend enthält die Matrix **B** alle die Informationen über die relevanten Beurteilungsdimensionen, d. h. die Skalen werden zu charakteristischen Beurteilungsdimensionen oder Basiseigenschaften der analysierten Produktkategorie zusammengefaßt. Die einzelnen Spalten der Komponentenmatrix **C** korrespondieren mit charakteristischen Konsumenten oder Konsumententypen bzw. Konsumentensegmenten.

Nachdem die inhaltliche Bedeutung der drei Komponentenmatrizen **A**, **B** und **C** diskutiert worden ist, muß die Bedeutung der Kernmatrix **G** spezifiziert werden. Ihre Dimensionalität entspricht der des ursprünglichen Datenkörpers. Die Anzahl der Komponenten der Matrizen **A**, **B** und **C** bestimmt dabei die Größe der Kernmatrix als eine  $(P \times Q \times R)$ -dimensionale Matrix. Hieraus wird unmittelbar ersichtlich, daß jede Spalte einer Komponentenmatrix mit einer Entität der Kernmatrix zusammenhängt. Das Element der Kernmatrix  $g_{pqr}$  mißt demzufolge den Zusammenhang zwischen der  $p$ -ten Komponente der Matrix **A**, der  $q$ -ten Komponente der Matrix **B** und der  $r$ -ten Komponente der Matrix **C**. Die Interpretation der Kernmatrix vereinfacht sich, wenn man berücksichtigt, daß die Summe der quadrierten Kernmatrixelemente gleich der erklärten Varianz (SS(Fit)) in den Beobachtungen ist (vgl. Kroonenberg 1983, S. 35). Der Wert  $g_{pqr}^2$  gibt demzufolge an, welchen Beitrag die Kombination der  $p$ -ten Komponente des ersten Modus mit der  $q$ -ten Komponente des zweiten Modus und der  $r$ -ten Komponente des dritten Modus an der gesamten Varianzerklärung des Modells hat. Die Interpretation der quadrierten Kernmatrixelemente als Varianzerklärungsanteile gilt jedoch ausschließlich bei

orthonormal beschränkten Komponentenmatrizen  $\mathbf{A}$ ,  $\mathbf{B}$  und  $\mathbf{C}$ .

In Matrixschreibweise stellt sich die Grundgleichung des Tucker3-Modells für die  $k$ -te Scheibe der Datenmatrix wie folgt dar:

$$\mathbf{X}_k = \mathbf{A} \sum_{r=1}^R c_{kr} \mathbf{G}_r \mathbf{B}'. \quad (3)$$

Nach Tucker erhält man die Komponentenmatrizen  $\mathbf{A}$ ,  $\mathbf{B}$  und  $\mathbf{C}$  über eine Singulärwertzerlegung der folgenden Produktmomentmatrizen.

$$\mathbf{M}_I = [m_{ii'}], \quad m_{ii'} = \sum_j \sum_k x_{ijk} x_{i'jk} \quad (4)$$

$$\mathbf{M}_J = [m_{jj'}], \quad m_{jj'} = \sum_i \sum_k x_{ijk} x_{ij'k} \quad (5)$$

$$\mathbf{M}_K = [m_{kk'}], \quad m_{kk'} = \sum_i \sum_j x_{ijk} x_{ij'k'} \quad (6)$$

Die Komponentenmatrizen  $\mathbf{A}$ ,  $\mathbf{B}$  und  $\mathbf{C}$  werden mit Hilfe einer Singulärwertzerlegung der Produktmomentmatrizen bestimmt, wobei lediglich die ersten  $P$ ,  $Q$  bzw.  $R$  Komponenten der entsprechenden Eigenvektoren erhalten bleiben:

$$\mathbf{M}_I = \mathbf{A} \mathbf{\Lambda}_i \mathbf{A}' \quad (7)$$

$$\mathbf{M}_J = \mathbf{B} \mathbf{\Lambda}_j \mathbf{B}' \quad (8)$$

$$\mathbf{M}_K = \mathbf{C} \mathbf{\Lambda}_k \mathbf{C}'. \quad (9)$$

Die Festlegung der relevanten Komponenten für die Lösungsmatrizen  $\mathbf{A}$ ,  $\mathbf{B}$  und  $\mathbf{C}$  kann anhand der Singulärwerte erfolgen, die auf der Diagonalen der Diagonalmatrizen  $\mathbf{\Lambda}_i$ ,  $\mathbf{\Lambda}_j$  und  $\mathbf{\Lambda}_k$  stehen. Tucker (1966) empfiehlt in diesem Zusammenhang einen Scree-Test der Singulärwerte. Die Schätzung der Kernmatrix  $\mathbf{G}$  ergibt sich dann über

$$g_{pqr} = \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K a_{ip} b_{jq} c_{kr} x_{ijk}. \quad (10)$$

Die Lösung nach Tucker (1966) ist allerdings bei einer nicht vollständigen Dekomposition der Daten in alle möglichen Komponenten  $I$ ,  $J$  und  $K$ , nicht optimal im Sinne der Minimierung eines Kleinst-Quadrat-Kriteriums. Daher haben Kroonenberg und de Leeuw (1980) einen alternierenden Kleinst-Quadrat-Algorithmus entwickelt, der die TIUCKALS3-Verlustfunktion minimiert. Der TUCKALS3-Algorithmus ergibt sich durch

$$\text{TUCKALS3}(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{G}_1, \dots, \mathbf{G}_R) = \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K \left( x_{ijk} - \sum_{p=1}^P \sum_{q=1}^Q \sum_{r=1}^R a_{ip} b_{jq} c_{kr} g_{pqr} \right)^2 \quad (11)$$

$$= \sum_{k=1}^K \left\| \mathbf{X}_k - \mathbf{A} \sum_{r=1}^R c_{kr} \mathbf{G}_r \mathbf{B}' \right\|^2 \quad (12)$$

mit  $\|\cdot\|^2$  als euklidische Norm einer Matrix. Dieser Algorithmus konvergiert im allgemeinen in das globale Minimum der Verlustfunktion (vgl. zu den Konvergenzeigenschaften Kroonenberg & de Leeuw 1980). Die Ähnlichkeit des TUCKALS3-Algorithmus mit dem ursprünglichen Tucker-Ansatz zeigt sich darin, daß in einer iterativen Prozedur Eigenvektoren und Eigenwerte geschätzt werden (Bauer-Rutishauser Methode). Aufgrund der Nähe des TUCKALS3-Algorithmus zur ursprünglichen Tucker-Methode lassen sich die nicht optimalen Parametermatrizen einer Tucker-Lösung zur Initialisierung des TUCKALS3-Algorithmus verwenden. Kroonenberg, ten Berge, Brouwer & Kiers (1989) haben schließlich den Algorithmus dahingehend verbessert, daß die Bauer-Rutishauser Methode durch eine nicht iterative Gram-Schmidt Orthogonalisierung verbessert wurde. Dieses Vorgehen ist in der aktuellen TUCKALS3 Version im Programmsystem 3WAYPACK Version 2.0 von Kroonenberg (1996) implementiert. Die Festlegung der Anzahl, der als relevant erachteten Komponenten ist jetzt auch nicht mehr, wie im Sinne Tuckers (1966), über einen Scree Test der Singulärwerte der Produktmomentmatrizen (7) - (9) möglich, sondern orientiert sich an den erklärten Varianzbeiträgen der einzelnen Komponentenmatrizen und am Gesamtfit des Modells.

Da das Modell unteridentifiziert ist (die Komponentenmatrizen können mit beliebigen nicht-singulären Matrizen transformiert werden), muß die Verlustfunktion beschränkt werden, um eine Lösung zu erhalten. Im allgemeinen restringiert man die Lösungsmatrizen **A**, **B** und **C** während der Schätzung auf spaltenweise Orthonormalität. Nachdem der Algorithmus konvergiert ist, kann die Beschränkung der Parametermatrizen aufgehoben werden. Die Matrizen können jeder beliebigen nicht-singulären Transformation unterzogen werden, sofern die inversen Transformationen in der Kernmatrix "aufgefangen" werden. Kürzlich haben Kiers, ten Berge und Kroonenberg (1992) einen modifizierten Algorithmus zur Schätzung des Tucker3-Modells vorgestellt. Dieser Algorithmus verzichtet auf die explizite Schätzung einer der Parametermatrizen (z. B. im Modus *A*). Dadurch wird es möglich, Datenmatrizen mit einer sehr großen Anzahl von Fällen in einem Modus (hier *A*) zu schätzen. Darüber hinaus ist dieser Algorithmus auch in der Lage Multitrait-Multimethod Matrizen zu bearbeiten.

### 3.3 Das Tucker2-Modell

Eine wichtige Vereinfachung des Tucker3-Modells ergibt sich, wenn man eine Komponentenmatrix (z. B. die Matrix **C**) als Identitätsmatrix annimmt. Das bedeutet, daß der dritte Modus nicht in seine Basisdimensionen zerlegt wird, sondern elementweise erhalten bleibt. Entsprechend dieser Spezifikation liegt die Kernmatrix jetzt als erweiterte Kernmatrix vor, die ebenso viele Scheiben/Lagen hat wie die dritte Modalität an Ausprägungen aufweist. Diese diskutierte Modellstruktur schlagen mehrere Autoren vor (u. a. Israelsson 1969, Jennrich 1972, Carroll & Chang 1970, Kroonenberg & de Leeuw 1977). Algebraisch ergibt sich dieses Modell, auch Tucker2-Modell genannt, durch



$$x_{ijk} = \hat{x}_{ijk} + e_{ijk} \quad (13)$$

$$\hat{x}_{ijk} = \sum_p \sum_q a_{ip} b_{jq} g_{pqk}, \quad (14)$$

bzw. in Matrixnotation

$$\hat{\mathbf{X}}_k = \mathbf{A} \mathbf{G}_k \mathbf{B}'. \quad (15)$$

Die Kernmatrixscheibe  $\mathbf{G}_k$  enthält die Information, wie die Beziehungen zwischen idealisierten Produkten (Matrix  $\mathbf{A}$ ) und charakteristischen Produkteigenschaften (Matrix  $\mathbf{B}$ ) für die  $k$ -te Person ausgestaltet sind. Kroonenberg und de Leeuw (1977, 1980) haben einen alternierenden Kleinste-Quadrate-Algorithmus für das Tucker2-Modell entwickelt, der in dem Programm TUCKALS2 des Programmsystems 3WAYPACK von Kroonenberg (1996) implementiert ist. Das Programm minimiert die TUCKALS2-Verlustfunktion im Hinblick auf das Kleinste-Quadrate-Kriterium.

$$\text{TUCKALS2}(\mathbf{A}, \mathbf{B}, \mathbf{G}_1, \dots, \mathbf{G}_K) = \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K \left( x_{ijk} - \sum_{p=1}^P \sum_{q=1}^Q a_{ip} b_{jq} g_{pqk} \right)^2 \quad (16)$$

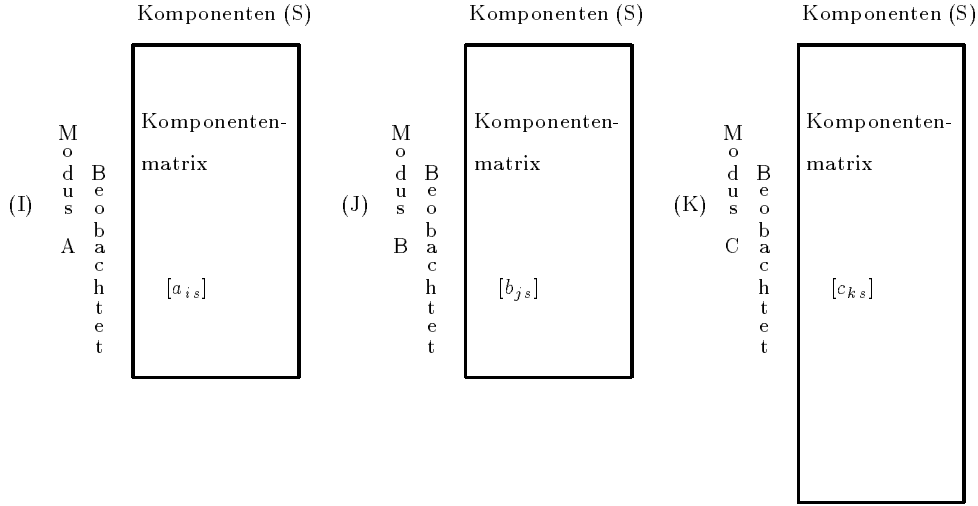
$$= \sum_{k=1}^K \|\mathbf{X}_k - \mathbf{A} \mathbf{G}_k \mathbf{B}'\|^2, \quad (17)$$

Dieser Algorithmus konvergiert im allgemeinen in das globale Minimum der Verlustfunktion (vgl. zu den Konvergenzeigenschaften Kroonenberg & de Leeuw 1977, 1980). Vergleicht man das Tucker3-Modell mit dem Tucker2-Modell, so muß das erste Modell als beschränkte Version des zweiten Modells angesehen werden. Die Beschränkung liegt darin, daß die Kernscheiben der erweiterten Kernmatrix als lineare Kombinationen einer geringeren Anzahl von Kernscheiben (Tucker3) angesehen werden können.

### 3.4 Das CANDECOMP/PARAFAC-Modell

Die Tucker-Modelle haben in ihrer ursprünglichen Form das Prinzip der Singulärwertzerlegung auf den Fall dreimodaler Daten übertragen. Ein drittes Modell, das CANDECOMP/PARAFAC-Modell kann hingegen als eine Generalisierung der Singulärwertzerlegung für den dreimodalen Fall angesehen werden. Carroll & Chang (1970) sowie Harshman (1970) entwickelten unabhängig voneinander ein mathematisch identisches Modell zur Dekomponierung dreimodaler Daten. Carroll & Chang nannten ihr Modell CANDECOMP, was für CANonical DECOMPosition steht. Das gleichzeitig entwickelte PARAFAC (PARAllel FACTor analysis) entstand vor dem Hintergrund der Übertragung des Prinzips der "parallel proportional profiles" (Cattell 1944) in ein mathematisches Modell. Nach Cattell sollen Faktoren und Faktorwerte unterschiedlicher Meßzeitpunkte ihre Struktur beibehalten und sich nur durch proportionale Änderungen voneinander unterscheiden dürfen. Harshmans

Abbildung 3: Die Struktur des PARAFAC-Modells



Leistung bestand darin, dieses Prinzip in ein Datenmodell zu übertragen, in dessen ursprünglicher Version der dritte Modus durch unterschiedliche Situationen gebildet wird (Profildaten). Algebraisch kann das Modell wie folgt beschrieben werden

$$x_{ijk} = \hat{x}_{ijk} + e_{ijk} \quad (18)$$

$$\hat{x}_{ijk} = \sum_{r=1}^R a_{ir} b_{jr} c_{kr}, \quad (19)$$

bzw. in Matrixnotation

$$\hat{\mathbf{X}}_k = \mathbf{A} \mathbf{\Lambda}_k \mathbf{B}'. \quad (20)$$

Die Koeffizienten  $a_{is}$ ,  $b_{js}$  und  $c_{ks}$  sind Elemente der Komponentenmatrizen  $\mathbf{A}$ ,  $\mathbf{B}$  und  $\mathbf{C}$  mit der Dimension  $(I \times S)$ ,  $(J \times S)$  und  $(K \times S)$ . Die Matrix  $\mathbf{\Lambda}_k$  ist die Diagonalmatrix der  $k$ -ten Zeile der Matrix  $\mathbf{C}$ . Die Komponentenmatrix  $\mathbf{A}$  enthält - vergleichbar der Interpretation der Tucker-Modelle - die Markentypen. Matrix  $\mathbf{B}$  repräsentiert die charakteristischen Produkteigenschaften, wohingegen die Matrix  $\mathbf{C}$  Konsumentensegmente angibt.

Die Proportionalität im Modell zeigt sich darin, daß die Elemente  $a_{ir}$ ,  $b_{jr}$  und  $c_{kr}$  Komponentenkoordinaten der Produkte, Skalen und Personen auf der  $r$ -ten CANDECOMP/PARAFAC-Komponente sind. Entsprechend sind nur proportionale Unterschiede zwischen den Produkten, Skalen und Personen möglich. Aufgrund dessen muß jeder Modus auf die gleiche Anzahl Komponenten reduziert werden. Im Gegensatz zu dem Tucker3-Modell entfällt die Schätzung einer Kernmatrix. Dies vereinfacht die Interpretation der Lösung, da keine Beziehungen zwischen den Komponenten der Modi zugelassen

sind. Das CANDECOMP/PARAFAC-Modell stellt lediglich einen Satz von Komponenten bereit, der über alle Modi hinweg gilt (z. B. Komponente 1 der Modi A, B, C). Das CANDECOMP/PARAFAC-Modell kann deshalb auch als ein Ansatz interpretiert werden, der gleichzeitig drei Modi in einem eindimensionalen Raum abbildet. Ein weiteres Charakteristikum ist die Unmöglichkeit der Rotation der Ergebnismatrizen ohne gleichzeitige Reduktion der Varianzerklärung. Damit unterscheidet sich CANDECOMP/PARAFAC nicht nur von den Tucker-Modellen, sondern auch von den klassischen faktorenanalytischen Modellen für den zweimodalen Fall. Die Parametermatrizen werden über die Minimierung der Verlustfunktion

$$\text{CP}(\mathbf{A}, \mathbf{B}, \mathbf{\Lambda}_1, \dots, \mathbf{\Lambda}_K) = \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K \left( x_{ijk} - \sum_{r=1}^R a_{ir} b_{jr} c_{kr} \right)^2 \quad (21)$$

$$= \sum_{k=1}^K \|\mathbf{X}_k - \mathbf{A} \mathbf{\Lambda}_k \mathbf{B}'\|^2, \quad (22)$$

mittels eines alternierenden Kleinste-Quadrate-Ansatzes ermittelt (vgl. Harshman 1970, S. 30-33), der sukzessive die nicht erklärte Varianz bis zur Erreichung eines Konvergenzkriteriums minimiert.

Eine methodische Weiterentwicklung kann in dem Ansatz PFCORE gesehen werden (Lundy, Harshman & Kruskal 1989). PFCORE erweitert den CANDECOMP/ PARAFAC-Ansatz um eine Kernmatrix. In der Regel baut PFCORE auf einer orthogonal beschränkten CANDECOMP/PARAFAC-Lösung auf. In der zweiten Stufe wird dann eine Kernmatrix mittels generalisierter invertierter Matrizen  $\mathbf{A}$ ,  $\mathbf{B}$  und  $\mathbf{C}$  bestimmt. Strenggenommen ist PFCORE damit nur eine andere Methode zur Schätzung des Tucker3-Modells. Das zweistufige Vorgehen impliziert aber, daß die Lösung dieses Vorgehens suboptimal ist, das heißt, daß die PFCORE Lösung niemals besser sein kann als eine Tucker3 Lösung, da die Bestimmung der Kernmatrix im PFCORE Ansatz nicht Gegenstand der Minimierung der Verlustfunktion ist, sondern erst auf einer nachgelagerten Stufe bestimmt wird.

### 3.5 Die Modellbewertung und Interpretation der dreimodalen Analyse

In diesem Kapitel werden verschiedene Probleme aufgegriffen, die bei der Anwendung der dreimodalen Analyse von Bedeutung sind. Die Beurteilung der Qualität einer Lösung, die Skalierung der Inputdaten, sowie die Behandlung fehlender Werte betrifft alle Modelle. Zusätzlich werden für die Tucker-Modelle einige ausgewählte Interpretationshilfen vorgestellt, die Konstruktion sogenannter Joint Plots und die Rotation der Ergebnismatrizen.

### 3.5.1 Die Beurteilung der Qualität einer Lösung

Die Gesamtvarianz der vorgestellten Modelle kann in die erklärte und die nicht-erklärte Varianz zerlegt werden:

$$\sum_i \sum_j \sum_k x_{ijk}^2 = \sum_i \sum_j \sum_k \hat{x}_{ijk}^2 + \sum_i \sum_j \sum_k e_{ijk}^2 \quad (23)$$

$$SS(\text{Total}) = SS(\text{Fit}) + SS(\text{Res}). \quad (24)$$

Die Qualität der Lösung ergibt sich dann unmittelbar aus der Relation von erklärter Varianz zur Gesamtvarianz. Im Falle der Tucker-Modelle läßt sich diese Bedingung darüber hinaus auch auf einzelne Elemente eines Modus übertragen (vgl. ten Berge, de Leeuw & Kroonenberg 1987). Die erklärte Varianz der  $t$ -ten Woche der Datenmatrix  $\mathbf{X}$  kann man wie folgt partitionieren:

$$\|\mathbf{X}_k\|^2 = \|\hat{\mathbf{X}}_k\|^2 + \|\mathbf{X}_k - \hat{\mathbf{X}}_k\|^2. \quad (25)$$

Diese Berechnungen ermöglichen nunmehr eine differenzierte Überprüfung des Fits für jeden Level eines Modus. Es wird damit ersichtlich, welche Elemente einen hohen Beitrag zur Erklärung des Modells leisten und welche nicht, oder anders ausgedrückt, welche Elemente gut bzw. schlecht durch das Modell repräsentiert werden.

Damit ist es möglich, in jedem der vorgestellten Modelle den jeweiligen Fit über den Anteil der erklärten Varianz an der Gesamtvarianz zu erklären.

$$\text{Fit} = \frac{SS(\text{Fit})}{SS(\text{Total})} \quad (26)$$

### 3.5.2 Die Datenvorverarbeitung

Im Gegensatz zu der Anwendung der klassischen zweimodalen Faktoren- oder Hauptkomponentenanalyse, die in der Regel auf Korrelationsmatrizen beruhen und damit automatisch mit standardisierten Daten arbeiten, entfällt dieser Schritt der Standardisierung der Daten bei den Anwendungen der dreimodalen Methoden. Den Grundgleichungen der Tucker3-, Tucker2- und CANDECOMP/PARAFAC-Modelle kann entnommen werden, daß die Modelle eine Reduktion der Dimensionalität in den Daten durch einen direkten Fit des Modells auf die Daten erreichen. Implizit setzen diese Modelle jedoch voraus, daß die Daten vorher einem gewissen Preprocessing unterzogen wurden. Unter Preprocessing versteht man die Zentrierung und/oder die Standardisierung eines Modus oder einer Kombination von Modi, um häufig auftretende Verteilungsprobleme bei Intervalldaten wie Schiefe oder Zweigipfeligkeit abzumildern. Die Schwierigkeiten des Preprocessing liegen vor allem in dem Komplexitätszuwachs durch die simultane Betrachtung dreier Modi. Jede Standardisierung der Daten in einem Modus hat Konsequenzen auf die Datenstruktur in den anderen Modi. Die Entscheidung für eine bestimmte Art des Preprocessing hängt von dem jeweiligen Datentyp ab, d.h. Profildaten sind anders zu behandeln, als Daten

die mit dem semantischen Differential erhoben wurden. Eine umfassende Diskussion des Datenpreprocessing für das CANDECOMP/PARAFAC-Modell findet sich bei Harshman und Lundy (1984) und soll an dieser Stelle nicht geführt werden. Eine Diskussion dieser Problematik für die Tucker-Modelle liefert Kroonenberg (1983, S. 127-157).

### 3.5.3 Die Behandlung fehlender Werte

Die hier vorgestellten dreimodalen Verfahren setzen eine vollständige Datenmatrix voraus, die keine fehlenden Werte enthält. Die Behandlung von fehlenden Werten ist kompliziert und erst wenige Analyseergebnisse zu dieser Problematik liegen vor. Das TUCKALS3 Programmsystem (Kroonenberg & Brouwer 1993) ist seit kurzem mit einer "missing value" Option ausgestattet. Die fehlenden Werte werden in einem iterativen Prozeß während des allgemeinen alternierenden Kleinste-Quadrate-Ansatzes geschätzt. Die Existenz von fehlenden Werten führt in der Regel zu einem Anstieg der erklärten Quadratsumme, da sowohl das Modell an die Daten, als auch die Daten an das Modell angepaßt werden (Kroonenberg & Brouwer 1993, S. 15). Diese Tatsache macht sich insbesondere bei einer großen Anzahl fehlender Werte bemerkbar. Es sollte deshalb der Anteil fehlender Werte möglichst niedrig sein, um den oben beschriebenen Einfluß auf die Lösung so gering wie möglich zu halten.

### 3.5.4 Die Konstruktion von Joint Plots

Die Option zur Konstruktion von Joint Plots steht nur dem Anwender der Tucker-Modelle zur Verfügung. Die methodische Darstellung wird hier für das Tucker3-Modell geführt, sie läßt sich aber auf das Tucker2-Modell übertragen und gilt dann entweder für den Fall der durchschnittlichen Kernmatrix oder eine spezifische Kernmatrixscheibe. Jede Scheibe der Kernmatrix ( $\mathbf{G}_r$ ) im Tucker3-Modell enthält die Informationen darüber, wie die Beziehungen zwischen den idealisierten Produkten (Komponenten der Matrix  $\mathbf{A}$ ) und den charakteristischen Eigenschaften (Komponenten der Matrix  $\beta f B$ ) ausgestaltet sind. Um dann die Beziehung zwischen den ursprünglich bewerteten Produkten und Skalen zu untersuchen, wird die Kernmatrixscheibe dekomponiert, da sie die Informationen über die Beziehungen der Komponenten der Matrizen  $\mathbf{A}$  und  $\mathbf{B}$  enthält (Kroonenberg 1985). Anschließend skaliert man die Komponentenmatrizen  $\mathbf{A}$  und  $\mathbf{B}$  mit Teillösungen der Singularwertzerlegung, um eine möglichst gute Repräsentation der Elemente der Modi A und B in einem gemeinsamen Raum zu erzielen.

$$\mathbf{H}_r = \mathbf{A}\mathbf{G}_r\mathbf{B}' \quad (27)$$

$$= \mathbf{A}(\mathbf{U}_r\mathbf{D}_r\mathbf{V}_r')\mathbf{B}' \quad (28)$$

$$= \left(\frac{I}{J}\right)^{\frac{1}{4}}(\mathbf{A}\mathbf{U}_r\mathbf{D}_r^{\frac{1}{2}}) \cdot \left(\frac{J}{I}\right)^{\frac{1}{4}}(\mathbf{B}\mathbf{V}_r\mathbf{D}_r^{\frac{1}{2}}) \quad (29)$$

$$= \tilde{\mathbf{A}}_r\tilde{\mathbf{B}}_r \quad (30)$$

$$\tilde{\mathbf{A}}_r = \left(\frac{I}{J}\right)^{\frac{1}{4}}(\mathbf{A}\mathbf{U}_r\mathbf{D}_r^{\frac{1}{2}}) \quad (31)$$

$$\tilde{\mathbf{B}}_r = \left(\frac{J}{I}\right)^{\frac{1}{4}} (\mathbf{B} \mathbf{V}_r \mathbf{D}_r^{\frac{1}{2}}) \quad (32)$$

$\tilde{\mathbf{A}}_r$  und  $\tilde{\mathbf{B}}_r$  sind die Koordinatenmatrizen der Modi  $A$  und  $B$  für die  $r$ -te Komponente der Matrix  $\mathbf{C}$  (hier Segment  $r$ ). Sie ermöglichen den gleichzeitigen Plot von Produkten und Skalen in einen Raum, der die Beurteilungen von Produkten und Eigenschaften dieses Konsumentensegments repräsentiert. Auf diese Weise ist es möglich, eine Verdichtung der Ergebnisse zu erreichen und die Interpretationsschwierigkeiten der Kernmatrix aufzulösen. Diese Methodik kann als Erweiterung der Biplot-Technik aufgefaßt werden (vgl. Gabriel 1971, 1981).

### 3.5.5 Rotationen

Eine weitere Möglichkeit, die Analyseergebnisse der Tucker-Modelle besser interpretieren zu können, besteht in der Rotation der Ergebnismatrizen. Rotationen gehören im klassischen Fall der zweimodalen Faktoren- oder Hauptkomponentenanalyse zu dem standardmäßigen Vorgehen. Auch im dreimodalen Fall können die Ergebnismatrizen bspw. auf eine Einfachstruktur rotiert werden. Allerdings müssen die inversen Rotationen in der Kernmatrix ausgeglichen werden. Es muß jedoch herausgestellt werden, daß die Rotation der Komponentenmatrizen die Interpretation der Kernmatrix erschweren kann. Die Rotation der Matrizen  $\mathbf{A}$ ,  $\mathbf{B}$  und  $\mathbf{C}$  auf Einfachstruktur korrespondiert nicht mit einer Einfachstruktur der Kernmatrix. Aus diesem Grunde schlägt Kroonenberg (1984, S. 77) vor, die Kernmatrix nach einem Zielkriterium zu rotieren. Sie weist dann eine Einfachstruktur auf, wenn die Nichtraumdiagonalelemente ( $p \neq q \neq r$ ) nahe Null sind und lediglich die Raumdiagonalelemente ( $p = q = r$ ) deutlich von Null abweichende Ladungen aufweisen.

## 4 Eine Anwendung der dreimodalen Analyse in der Marktforschung

### 4.1 Die Vorgehensweise bei der Datenanalyse

Die Anwendungspotentiale der Verfahren zur Analyse dreimodaler Daten sollen an einem empirischen Datensatz zur Beurteilung von Geschirrspülmitteln durch die Konsumenten verdeutlicht werden. Der Datensatz wurde im Rahmen einer studentischen Studie im Markt für Geschirrspülmittel erhoben. Er umfaßt die Beurteilung von sechs Produkten auf 11 Zustimmungsskalen vom Likert-Typ durch 137 Befragungspersonen. Als Produkte wurden die Marken "Pril", "Palmolive", "Fairy Ultra", "Frosch", "Pril Balsam" und "Palmolive Sensitive" beurteilt. Allerdings mußten die Befragungspersonen nur die Produkte Ihres Evoked-Set bewerten, für die eine Kaufwahrscheinlichkeit größer als Null vorliegt. Somit haben die meisten Personen weniger als die sechs vorgegebenen Produkte beurteilt. Insgesamt sind deshalb 45 Prozent der Datenpunkte (Produkte  $\times$  Skalen  $\times$  Befragungspersonen) missing values. Dieser Prozentsatz ist zwar relativ hoch, er erlaubt aber dennoch

die Bestimmung einer stabilen Lösung.

Der erste Schritt der Datenanalyse beinhaltet - unabhängig von der Wahl der Analyse-methode - die Entscheidung über das geeignete Datenpreprocessing. Im Rahmen dieses Schrittes sind mögliche Zentrierungen und Standardisierungen festzulegen. Bei dem hier vorliegenden Datensatz kommt entsprechend den Ausführungen Kroonenbergs (1983, Kapitel 6) eine Zentrierung der Skalen in Betracht. Die Zentrierung der Skalen verdeutlicht die Unterschiedlichkeit der Produkte auf den relevanten Beurteilungsdimensionen, allgemeine Produktähnlichkeiten oder -unähnlichkeiten bleiben jedoch in den Daten erhalten. Außerdem ist diese Art der Zentrierung besonders angebracht, wenn die Befragungspersonen jeweils ein Produkt vollständig oder gar nicht beurteilt haben. Die gewählte Zentrierungsprozedur steht darüber hinaus in Einklang mit der üblichen Standardisierung bei der Analyse von Korrelationsmatrizen durch die Hauptkomponentenanalyse oder Faktorenanalyse.

Im zweiten Schritt der Datenanalyse - der Parameterschätzung - wurde das Programm TUCKALS3 eingesetzt, das die umfassendste Modellierung, d. h. mit Beziehungsstruktur der Modi in der Kernmatrix ermöglicht. Nach einer Serie von unterschiedlichen Modellspezifikationen mit plausiblen Angaben über die Zahl von Komponenten pro Modus erwies sich eine  $3 \times 3 \times 3$  Komponentenlösung als adäquat zur Abbildung der Struktur in den Daten. Die Lösung war mit 61,06 Prozent erklärter Varianz akzeptabel und liefert interpretierbare Strukturen.

## 4.2 Die Ergebnisinterpretation der Komponentenlösung

Das Ergebnis der dreimodalen Datenanalyse mit jeweils drei Komponenten für jeden Modus gibt bereits auf der Komponentenebene Hinweise über die inhaltliche Struktur der Daten, wobei im ersten Modus (dem Produktmodus) je Komponente sechs Ladungen berechnet wurden, im zweiten Modus (dem Eigenschaftsmodus) je Komponente elf Ladungen und im dritten Modus (dem Personenmodus) sich so viele Ladungen pro Komponente ergaben, wie Personen in der Stichprobe vorhanden waren. Hinzu kommen die berechneten Koeffizienten aus der Kernmatrix, die aufgrund der vorliegenden  $3 \times 3 \times 3$  Komponentenlösung ebenfalls einen  $3 \times 3 \times 3$  Kubus von Koeffizienten bilden. Die Ladungsmatrizen **A**, **B** und **C** wurden jeweils mit der VARIMAX-Rotation auf Einfachstruktur rotiert. Da die Komponentenmatrizen orthonormal beschränkt in die Rotation eingehen, stimmt die rotierte Lösung mit einer ortho-obliquen Rotation nach Harris und Kaiser (1964) überein, die sich um Ladungsmatrizen niedriger Komplexität zu erzeugen als besonders geeignet bei der dreimodalen Analyse herausgestellt hat (Hakstian 1971, Kiers & ten Berge 1994).

Zur inhaltlichen Beurteilung des Ergebnisses werden hier zunächst die Komponentenstruktur des Produkt- und des Eigenschaftsmodus herangezogen, die eine Aussage über

Tabelle 1: Die Komponentenmatrix der Marken

| Komponentenmatrix <b>A</b> , VARIMAX-rotiert<br>Länge der Komponenten: $\mathbf{a}'\mathbf{a} = I$ |       |       |       |     |
|----------------------------------------------------------------------------------------------------|-------|-------|-------|-----|
| Label                                                                                              | $A_1$ | $A_2$ | $A_3$ | FIT |
| Pril                                                                                               | .06   | -.10  | 1.71  | .51 |
| Palmolive                                                                                          | 1.02  | -.18  | .67   | .59 |
| Fairy Ultra                                                                                        | -.13  | .13   | 1.60  | .72 |
| Frosch                                                                                             | .03   | 2.43  | .03   | .68 |
| Pril Balsam                                                                                        | 1.40  | .22   | -.13  | .57 |
| Palmolive Sensitive                                                                                | 1.72  | -.11  | -.23  | .63 |
| Bedeutung                                                                                          | .31   | .12   | .18   | .61 |

die vorliegende Marktstruktur zulassen. Die Komponentenladungsmatrizen sind in den Tabellen 1 und 2 zusammengestellt. Zur Interpretation sind die Ladungen auf die Anzahl der jeweils zugrundeliegenden Beobachtungen normiert worden, d. h. daß die berechneten Komponentenladungen mit einem Wert über 1.0 ein überdurchschnittliches Gewicht haben und geeignet sind, die Bedeutung der Komponente festzulegen.

Im Produktmodus wird deutlich, daß sich der Markt aus drei unterschiedlich bewerteten Produktgruppen zusammensetzt, wobei eine Produktgruppe (Komponente  $A_1$ ) durch Hautpflegeeigenschaften gebildet wird, die zweite Produktgruppe mit Frosch ( $A_2$ ) die Umweltfreundlichkeit erfaßt und die dritte Produktgruppe die eingeführten Standardprodukte kennzeichnet. Hier ist interessant, daß sich Palmolive - obwohl lange im Markt - offensichtlich auch in der Standardversion in das Pflegesegment positioniert hat.

Die wahrgenommene Struktur des Marktes wird durch die Ladungsmatrix für den Eigenschaftsmodus gestützt. Auch diese Ladungen sind über die Anzahl der Items normiert, so daß ein absoluter Wert von über 1.0 einen überdurchschnittlichen Einfluß indiziert. Die erste Komponente ( $B_1$ ) erfaßt die reinigungsbezogenen Eigenschaften eines Geschirrspülmittels (fettlösend, Glanz, Spülkraft, aber nicht unbedingt umweltgerecht), die zweite Komponente ( $B_2$ ) faßt die umweltrelevanten Eigenschaften (biologisch abbaubar, umweltschonend verpackt, aber ohne Trockenhilfe) zusammen und die dritte Komponente ( $B_3$ ) den Hautpflegeaspekt (händeschonend, pflegend, aber ohne Trockenhilfe).

Der dritte Modus ordnet die Personen über die Komponentenladungen den drei durch die Komponenten repräsentierten Segmenten zu. Zur Illustration werden die Komponentenladungen für jeweils drei besonders charakteristische Personen aus jedem Segment in Tabelle 3 wiedergegeben, die übrigen sind weggelassen. Die Ladungen machen zum einen deutlich, wie die Zuordnung zu den Segmenten erfolgt. Sie zeigen aber auch Zu-



Tabelle 2: Die Komponentenmatrix der Skalen

| Komponentenmatrix <b>B</b> , VARIMAX-rotiert       |       |       |       |     |
|----------------------------------------------------|-------|-------|-------|-----|
| Länge der Komponenten: $\mathbf{b}'\mathbf{b} = J$ |       |       |       |     |
| Label                                              | $B_1$ | $B_2$ | $B_3$ | FIT |
| Wirkt besonders gut fettlösend                     | -1.21 | -.24  | -.60  | .46 |
| Schont die Hände                                   | .55   | .32   | 1.36  | .33 |
| Sorgt für strahlenden Glanz                        | -1.10 | -.27  | -.10  | .44 |
| Hat eine besonders hohe Spülkraft                  | -1.20 | -.25  | -.29  | .55 |
| Enthält hautpflegenden Wirkstoff                   | .63   | .22   | 1.80  | .53 |
| Hat beim Spülen angenehmen Duft                    | .11   | .29   | 1.07  | .26 |
| Das Produkt ist biologisch abbaubar                | -.40  | -1.45 | -.70  | .49 |
| Macht Abtrocknen überflüssig                       | .79   | 2.37  | -1.70 | .89 |
| Ist besonders sparsam im Verbrauch                 | -.38  | .06   | -.16  | .07 |
| Entwickelt langanhaltenden Schaum                  | -.07  | .54   | .27   | .12 |
| Umweltgerechte Verpackung                          | 2.29  | -1.59 | -.94  | .91 |
| Bedeutung                                          | .26   | .19   | .15   | .61 |

Tabelle 3: Segmentspezifische Ausschnitte der Komponentenmatrix der Befragungspersonen

| Komponentenmatrix <b>C</b> , VARIMAX-rotiert       |          |          |          |          |
|----------------------------------------------------|----------|----------|----------|----------|
| Länge der Komponenten: $\mathbf{c}'\mathbf{c} = K$ |          |          |          |          |
| Label                                              | $C_1$    | $C_2$    | $C_3$    | FIT      |
| Vp003                                              | 2.54     | .07      | -.33     | .81      |
| VP043                                              | 2.48     | -.10     | .39      | .84      |
| Vp101                                              | 2.53     | -.02     | .25      | .83      |
|                                                    | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ |
| Vp017                                              | .04      | 2.48     | .97      | .61      |
| Vp037                                              | .19      | 2.47     | -.11     | .88      |
| Vp059                                              | -.01     | 2.90     | .45      | .90      |
|                                                    | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ |
| Vp007                                              | 1.96     | 1.37     | 4.64     | .78      |
| Vp018                                              | 1.98     | -.51     | 2.10     | .80      |
| Vp115                                              | .71      | .82      | 2.41     | .68      |
| Bedeutung                                          | .33      | .19      | .09      | .61      |

Tabelle 4: Die Kernmatrix

| Modus | $C_1$  |       |       | $C_2$ |        |       | $C_3$ |       |        |
|-------|--------|-------|-------|-------|--------|-------|-------|-------|--------|
|       | $B_1$  | $B_2$ | $B_3$ | $B_1$ | $B_2$  | $B_3$ | $B_1$ | $B_2$ | $B_3$  |
| $A_1$ | -28.36 | 16.20 | 22.86 | -5.21 | -24.32 | 18.58 | 12.86 | 9.61  | 16.46  |
| $A_2$ | -21.65 | 1.35  | .47   | -.50  | -18.52 | 1.20  | 4.46  | -2.95 | -17.74 |
| $A_3$ | -32.16 | 11.77 | 7.75  | -9.42 | -21.96 | 3.20  | 2.58  | 2.20  | .31    |

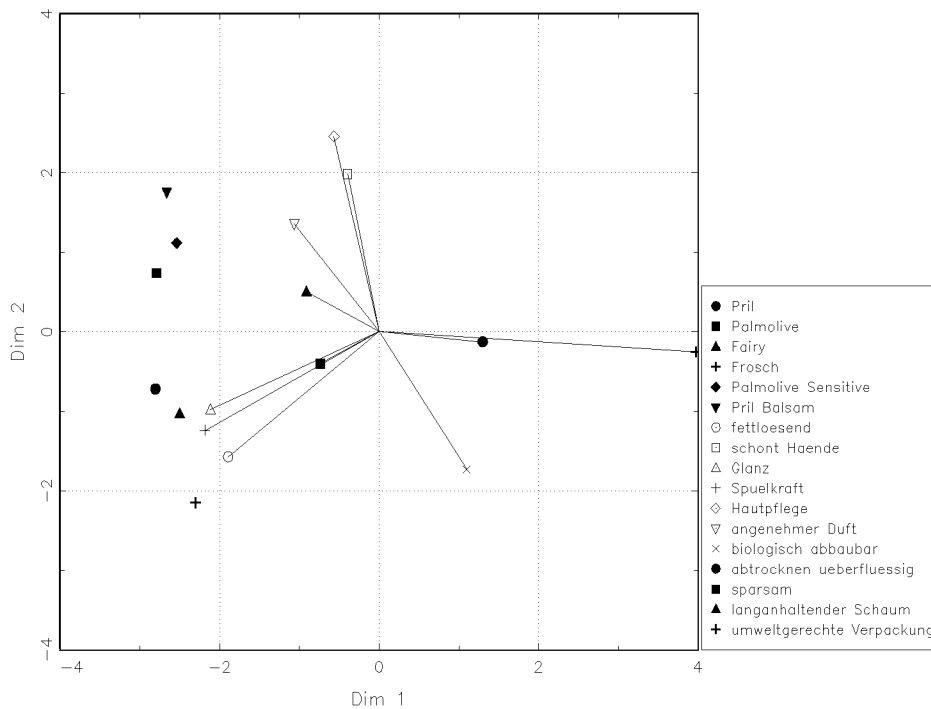
sammenhänge zwischen den Segmenten wie z. B. gleichzeitige hohe Ladungen in den Komponenten  $C_3$  und  $C_1$ .

Die Koeffizienten der Kernmatrix der dreimodalen Komponentenstrukturen sind in Tabelle 4 zusammengestellt. Jedes zweidimensionale Array bildet eine Scheibe der Kernmatrix ab. Obwohl die Interpretation äußerst schwierig ist, läßt sich aus den Zusammenhangskoeffizienten in der Kernmatrix einiges erkennen. Zunächst wird aus der Struktur deutlich, daß in den einzelnen Segmenten (Komponenten des Modus C) die Zusammenhänge zwischen den Komponenten in Modus A und B unterschiedlich bewertet werden. Zwar stimmen viele der Kernmatrixkoeffizienten eines Segments in ihrem Vorzeichen mit den Koeffizientenvorzeichen eines anderen Segments überein, die absolute Höhe der Kernmatrixkoeffizienten verdeutlicht jedoch Differenzen in der Produktwahrnehmung und -bewertung durch die entsprechenden Segmente. Das erste Konsumentensegment bescheinigt allen Produkttypen eine mangelhafte Verpackung im Hinblick auf die Umweltverträglichkeit sowie eine hohe Reinigungskraft. Erwartungsgemäß werden den hautpflegenden Produkten auch die Eigenschaften der Hautpflege bescheinigt. Dem Produkt Frosch aber wird in diesem Segment keine besondere Umweltverträglichkeit beigemessen. Das zweite Konsumentensegment bewertet alle Produkttypen als umweltverträglich und sieht darüber hinaus die hautpflegeintensiven Spülmittel auch als solche an. Das dritte Segment nimmt die Spülmittel, die sich als hautpflegend positioniert haben, auch als hautfreundlich wahr, es schreibt diesen Produkten aber auch eine mangelnde Reinigungskraft und fehlende Umweltverträglichkeit zu. Darüber hinaus wird das Spülmittel Frosch als besonders hautunfreundlich wahrgenommen.

### 4.3 Die Ergebnisinterpretation der Beurteilungsräume

Will man die Ähnlichkeiten und Unterschiede des wahrgenommenen Produktraums jetzt differenzierter sichtbar machen, kann über die Singulärwertzerlegung jeder Scheibe der Kernmatrix ein Joint Plot für jedes Segment erstellt werden. Es können auf diese Weise für jede Gruppe von Konsumenten die Beurteilungsräume mit den Produkten und Eigenschaften aufgespannt werden. Die Joint Plots der drei Segmente sind in den Abbildungen 4 bis 6 wiedergegeben. Die Skalenwerte der zugrundeliegenden Dimensionen orientieren sich in der Matrix an der Eigenschaft mit dem größten erklärten Varianzanteil. Die Di-

Abbildung 4: Joint Plot des 1. TUCKALS3-Käufersegments

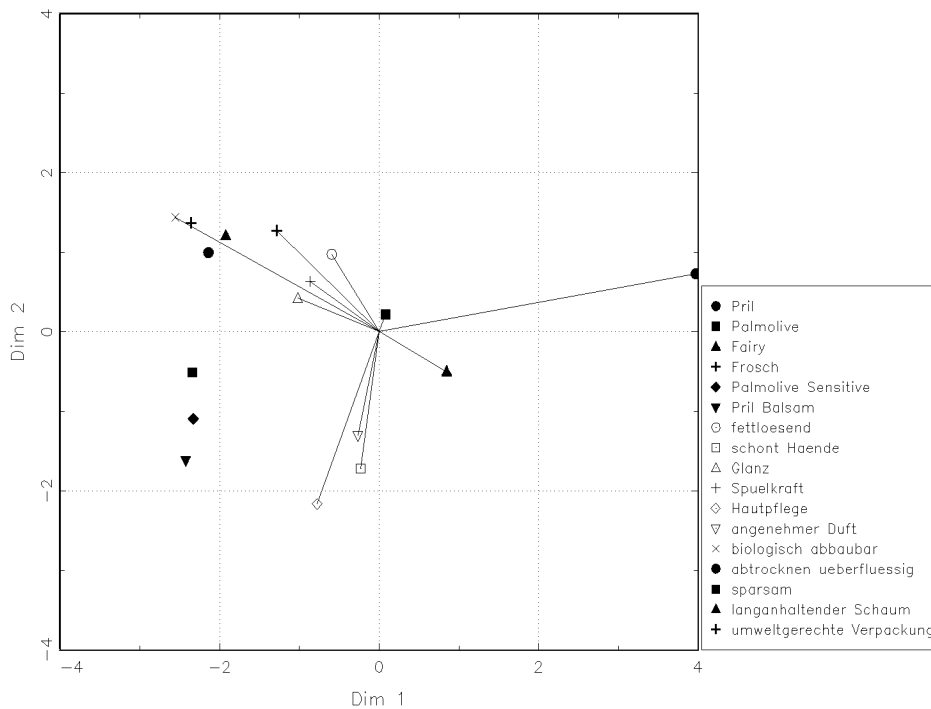


mensionen selbst haben keine inhaltliche Bedeutung. Nur die Ordnung und die Distanzen der Stimuli und Eigenschaften sind interpretierbar. Aus Gründen einer besseren Interpretation der Joint Plots sind im Plot des ersten Segments der Vektor der Eigenschaft "umweltgerechte Verpackung" und im Plot des zweiten Segments der Vektor der Eigenschaft "abtrocknen überflüssig" bis auf den Rand verkürzt dargestellt.

Die Eigenschaften sind mit einer Linie aus dem Ursprung und die Produkte mit einfachen Symbolen gekennzeichnet. Die Länge des Vektors bzw. der Abstand eines Symbols vom Ursprung des Koordinatenkreuzes ist ein Maß für die relative Bedeutung des korrespondierenden Produkts bzw. der Eigenschaft. Je länger der Vektor desto bedeutender ist die Ausprägung. Es wird zunächst beim Vergleich aller drei Plots in den Abbildungen 4 bis 6 deutlich, daß im Segment 3 und 1 die Produkte differenzierter wahrgenommen werden als im Segment zwei. Besonders prägnant ist dabei, daß die Hautpflege-Produkte aber in allen Segmenten als eine Gruppe von Produkten inklusive der Standardversion perzipiert werden. Ebenso liegen Fairy und Pril in allen Segmenten nahe beieinander. Alle Plots zeigen aber auch, daß die Standardversion von Palmolive immer am nächsten zu den anderen hautpflegenden Produkten liegt. Das Spülmittel Frosch dagegen wird nur im Vergleich zu den Standardprodukten differenziert.

Bezieht man bei der Interpretation die Urteile auf den Eigenschaften mit ein, so kann

Abbildung 5: Joint Plot des 2. TUCKALS3-Käufersegments

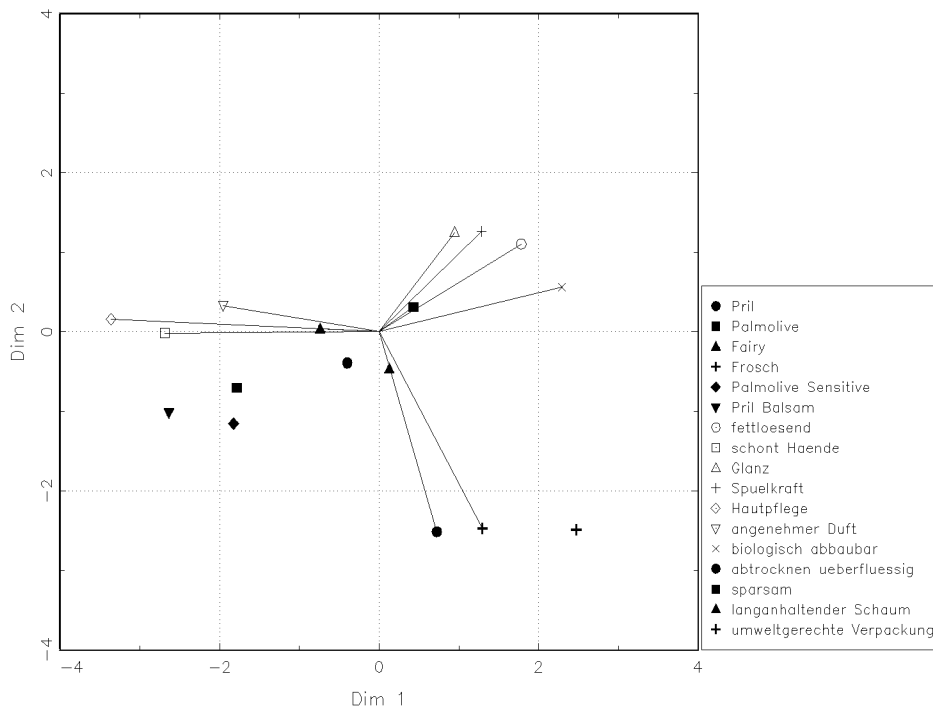


aus der Nähe und Richtung der Vektoren auf die differenzierenden Merkmale geschlossen werden. So liegen im Segment 1 primär wahrgenommene Unterschiede in der Hautpflegewirkung und Spülwirkung der Produkte vor. Ökologische Verpackung ist hier kein Differenzierungsmerkmal. Alle Produkte werden als nicht-ökologisch verpackt eingestuft. Auch im zweiten Segment ist die Hautpflegeeigenschaft ein wichtiges Differenzierungsmerkmal von den allgemeinen Eigenschaften wie Spülwirkung. Offensichtlich schreiben hier die Konsumenten den Standardprodukten etwas umweltfreundlichere Verpackungsmerkmale zu. Frosch wird zwar am deutlichsten mit biologischer Abbaubarkeit assoziiert, die Diskriminanzkraft ist aber äußerst gering. Die Segmenthaushalte sehen bei keinem der Produkte besondere Trockeneigenschaften.

Im dritten Segment bildet die Pflegeeigenschaft und die ökologische Verträglichkeit das Hauptdifferenzierungsmerkmal. Hier wird dem Spülmittel Frosch sogar die Fähigkeit der Einsparung des Abtrocknens zugewiesen, aber klar auch zwischen Standardspülmitteln und hautpflegenden Mitteln differenziert.

Die Analyse verdeutlicht, daß bei einer undifferenzierten Betrachtung des Marktes wichtige marketingrelevante Informationen über Teilsegmente verloren gegangen wären. Die dreimodale Analyse hat gleichzeitig die Beurteilung der Produkte des Marktes als Ausgangspunkt für die Segmentierung verwendet. Die Segmente könnten jetzt über ih-

Abbildung 6: Joint Plot des 3. TUCKALS3-Käufersegments



re soziodemographische Struktur identifiziert und differenziert bearbeitet werden bzw. Grundlage von Positionierungsentscheidungen sein.

## 5 Zusammenfassung

Der Beitrag führt in einen methodischen Ansatz ein, der in der Marktforschung bisher kaum Beachtung fand, was zum Teil allein an der Modellkomplexität und an der aufwendigen technischen Durchführung der dreimodalen Analyse lag. Mit der Existenz von leicht zu bedienender Software und einer leistungsfähigen PC-Generation steht heute der Anwendung der dreimodalen Datenanalyse aber nichts mehr im Wege. Die dreimodale Analyse hat für die Marktforscher den Vorteil, daß die Aufdeckung von Beurteilungsstrukturen und die Segmentierung nicht unabhängig voneinander durchgeführt werden, sondern der Interaktion zwischen Personen und den Beurteilungen von Stimuli direkt Rechnung getragen wird.

Die hier dargestellte Analyse mit dem Programm TUCKALS3 liefert dabei die umfassendste Form der dreimodalen Hauptkomponentenanalyse. Die Interpretation verdeutlicht die Fähigkeiten des heute existierenden Programms für explorative Analysen. Weitergehende Programmentwicklungen werden nicht nur das reine Datenverdichten und Segmentieren ermöglichen, sondern auch den Test von Strukturhypothesen im Sinne einer konfir-

matorischen Analyse zulassen. Die entsprechende Software liegt bereits in rudimentärer Form vor.

Die Potentiale der dreimodalen Analyse für die Marktforschung hängen aber weitgehend von den verfügbaren Informationen zur Lösung eines Marketingproblems ab. Häufig liefert die Analyse von Personenkomponenten wenig zusätzliche Information, oder die Zahl der Stimuli ist zu gering, so daß zweimodale Analysen, z. B. mit dem TUCKALS2 Programm, ausreichen. Hat man aber Daten über drei Modalitäten, wie sie in der Marktforschung meist vorzufinden sind, z. B. Produkt-/Werbekonzepte  $\times$  Skalen  $\times$  Personen oder Produkte  $\times$  Wahlindizes  $\times$  Wochen, so ist die dreimodale Analyse im vollständigen Modell der TUCKALS3-Analyse das Modell, das die umfassendste Information über die Marktstruktur für den Marketingmanager liefert.

## 6 Literatur

- Backhaus, K.; Erichson, B.; Plinke W.; Weiber, R. (1996): Multivariate Analysemethoden, Berlin 1993.
- Bentler, P. M.; Lee, S.-Y. (1978): Statistical aspects of a three-mode factor analysis model, in: Psychometrika, 43 (1978), S. 343-352.
- Bentler, P. M.; Poon, W.-Y.; Lee, S.-Y. (1988): Generalized multimode latent variable models: Implementation by standard programs, in: Computational Statistics and Data Analysis, 6 (1988), S. 107-118.
- Bloxom, B. (1968): A note on invariance in three-mode factor analysis, in: Psychometrika, 33 (1968), S. 347-350.
- Browne, M. W. (1984): The decomposition of multitrait-multimethod matrices, in: British Journal of Mathematical and Statistical Psychology, 37 (1984), S. 1-21.
- Carroll, J. D.; Chang, J. J. (1970): Analysis of individual differences in multidimensional scaling via an n-way generalization of "Eckart-Young" decomposition, in: Psychometrika, 35 (1970), S. 283-319.
- Cattell, R. B. (1944): "Parallel proportional profiles" and other principles for determining the choice of factors by rotation, in: Psychometrika, 9 (1944), S. 267-283.
- Eckart, C.; Young, G. (1936): The approximation of one matrix by another of lower rank, in: Psychometrika, 1 (1936), S. 211-218.
- Gabriel, K. R. (1971): The biplot graphic display of matrices with application to principal component analysis, in: Biometrika, 58 (1971), S. 453-467.

- Gabriel, K. R. (1981): Biplot display of multivariate matrices for inspection of data and diagnosis, in: Barnett, V. (Hrsg.): Interpreting multivariate data, Chicester 1981, S. 147-173.
- Hakstian, A. R. (1971): A comparative evaluation of several prominent methods of oblique factor transformation, in: Psychometrika, 36 (1971), S. 175-193.
- Harshman, R. A. (1970): Foundations of the PARAFAC procedure: Models and conditions for an "explanatory" multi-mode factor analysis, in: UCLA Working Papers in Phonetics, 16 (1970), S. 1-84.
- Harshman, R. A.; M. E. Lundy (1984): Data preprocessing and the extended PARAFAC model, in: Law, H. G.; Snyder Jr.; C. W.; Hattie, J. A.; McDonald R. P. (Hrsg.): Research methods for multimode data analysis. New York 1984, S. 216-284.
- Hildebrandt, L.; Klapper, D. (1994): The analysis of three-way three-mode data: A program based on GAUSS, in: Faulbaum, F. (Hrsg.), SoftStat '93: Advances in Statistical Software 4, Stuttgart 1994, S. 527-534.
- Israelsson, A. (1969): Three-way (or second order) component analysis, in: Wold, H.; Lyttkens, E. (Hrsg.): Nonlinear iterative partial least-squares (NIPALS) estimation procedures, in: Bulletin of the International Statistical Institute, 43 (1969), S. 29-51.
- Jennrich, R. A. (1972): A generalization of the multidimensional scaling model of Carroll & Chang, in : UCLA Working Paper in Phonetics, 22 (1972), S. 45-47.
- Kiers, H. A. L.; Kroonenberg, P. M.; ten Berge, J. M. F. (1992): An efficient algorithm for TUCKALS-3 on data with large numbers of observation units, in: Psychometrika, 57 (1992), S. 415-422.
- Kiers, H. A. L.; ten Berge, J. M. F. (1994): The Harris-Kaiser independent cluster rotation as a method for rotation to simple component weights, in: Psychometrika, 59 (1994), S.81-90.
- Kroonenberg, P. M. (1983): Three-mode principal component analysis: Theory and applications, Leiden 1983.
- Kroonenberg, P. M. (1984): Three-mode principal component analysis: Illustrated with an example from attachment theory, in: Law, H. G.; Snyder Jr.; C. W.; Hattie, J. A.; McDonald R. P. (Hrsg.): Research methods for multimode data analysis. New York (1984), S. 64-121.
- Kroonenberg, P. M. (1985): Three-mode principal components analysis of semantic differential data: The case of a triple personality, in: Applied Psychological Measurement, 9 (1985), S. 83-94.

- Kroonenberg, P. M. (1996): 3WAYPACK user's manual, version 2, Rijks Universiteit Leiden 1996.
- Kroonenberg, P. M.; Brouwer, P. (1993): TUCKALS3 user's manual. Rijks Universiteit Leiden 1993.
- Kroonenberg, P. M.; de Leeuw, J. (1977): A principal component analysis of three-mode data. Leiden: Department of Data Theory, University of Leiden 1977.
- Kroonenberg, P. M.; de Leeuw, J. (1980): Principal component analysis of three-mode data by means of alternating least squares algorithms, in: *Psychometrika*, 45 (1980), S. 69-97.
- Kroonenberg, P. M.; ten Berge, J. M. F.; Brouwer, P.; Kiers, H. A. L. (1989): Gram-Schmidt versus Bauer-Rutishauser in alternating least-squares algorithm for three-mode principal component analysis, in: *Computational Statistics Quarterly*, 2 (1989), S. 81-87.
- Lohmöller, J.-B. (1984): LVPLS program manual version 1.6, in: *Zentralarchiv für empirische Sozialforschung*, Universität zu Köln 1984.
- Lundy, M. E.; Harshman, R. A.; Kruskal, J. B. (1989), A two-stage procedure incorporating good features of both trilinear and quadrilinear models, in: Coppi, R.; Bolasco S. (Hrsg.): *Multiway data analysis*. Amsterdam 1989, S. 123-130.
- McDonald, R. P. (1978): A simple comprehensive model for the analysis of covariance structures, in: *British Journal of Mathematical and Statistical Psychology*, 22 (1978), S. 149-163.
- ten Berge, J. M. F.; de Leeuw, J.; Kroonenberg, P. M. (1987): Some additional results on principal components analysis of three-mode data by means of alternating least squares algorithms, in: *Psychometrika*, 52 (1987), S. 183-191.
- Tucker, L. R. (1963): Implications of factor analysis of three-way matrices for measurement of change, in: Harris, C. W. (Hrsg.): *Problems in Measuring Change*, Madison 1963, S. 122-137.
- Tucker, L. R. (1964): The extension of factor analysis to three-dimensional matrices, in: Frederiksen, N.; Gulliksen, H. (Hrsg.): *Contributions to mathematical psychology*, New York 1964, S. 110-127.
- Tucker, L. R. (1966): Some mathematical notes on three-mode factor analysis, in: *Psychometrika*, 31 (1966), S. 279-311.
- Tucker, L. R. (1972): Relations between multidimensional scaling and three-mode factor analysis, in: *Psychometrika*, 37 (1972), S. 3-27.