

Titel des Seminars: Einführung in die Skalierung und Clusteranalyse (V/Ue) (09295)  
Studienfach: Allgemeine Soziologie  
Teilprüfung in: Seminar  
Semester: 6 / 6  
DozentIn: Prof. Bacher

## **explorative und konfirmatorische Faktorenanalyse (Einführung in die Verfahren und Fallstudie anhand von Symlog-Daten)**

eingereicht am  
Lehrstuhl für Soziologie  
(Prof. Bacher)

eingereicht als: Allgemeine Soziologie und sozialwissenschaftliche Methoden einschließlich ihrer  
Anwendung in empirischen Untersuchungen

Verfasser/in:	Neubarth, Wolfgang	Ringleb, Stefan
Strasse:	Torwartstrasse 20	Herschlestrasse 17
PLZ Ort:	90480 Nürnberg	90443 Nürnberg
Telefon:	0911 – 40 10 294	0911 – 42 95 80
e-Mail:	wolfgang@neubarth.net	Stringleb@yahoo.de

Abgabetermin: 15.05.2000

Anzahl der Wörter/Zeichen: 62.581

## Inhaltsverzeichnis

<b>1.</b>	<b>Einleitung</b>	<b>3</b>
<b>2.</b>	<b>Explorative Faktorenanalyse</b>	<b>3</b>
2.2	Kommunalitäten	4
2.3	Faktorenextraktion	5
2.3.1	Überblick	6
2.3.2	Grundlegende Konzepte	7
2.4	Rotation	9
2.5	Faktorenwerte	10
<b>3.</b>	<b>Anwendung des Verfahrens</b>	<b>10</b>
3.1	Analyse der dimensionsbezeichnenden Items	12
3.2	Erweiterung auf 18 Items	14
3.2.1	4 Faktorenlösung	15
3.2.2	3 Faktorenlösung	16
3.3	Analyse des gesamten Symlog-Raums (alle 26 Items)	17
3.4	Zusammenfassung	19
<b>4.</b>	<b>konfirmatorische Faktorenanalyse</b>	<b>19</b>
4.1	Kurze Einführung	20
4.1.1	Grafische Steuerung	20
4.1.2	Kommando Steuerung	21
4.2	Identifikation	22
<b>5.</b>	<b>Der konkrete Anwendungsfall</b>	<b>24</b>
5.1	Möglichkeit: Neues Modell erstellen	26
5.2	Möglichkeit: Zusätzliche Angaben	27
5.3	Ebenenkonstukt	29
<b>6.</b>	<b>Diskussion</b>	<b>32</b>
<b>7.</b>	<b>Fazit</b>	<b>33</b>
<b>8.</b>	<b>Literatur</b>	<b>34</b>
<b>9.</b>	<b>Abbildungsverzeichnis</b>	<b>35</b>
<b>10.</b>	<b>Tabellenverzeichnis</b>	<b>35</b>
<b>11.</b>	<b>Anhang</b>	<b>36</b>

## 1. Einleitung

Ursprünglich wurde die Faktorenanalyse entwickelt um die menschliche Intelligenz besser erfassen zu können. Bereits 1904 schrieb Spearman seinen Aufsatz „General Intelligence“, objectively determined and measured“. Er stellte darin das erste faktorenanalytische Modell vor (Pawlik, 1976).

Über den Zeitraum von hundert Jahren hat sich nicht nur das erste Modell von Spearman verbessert. Der Anwendungsbereich ist, aufgrund der Stabilität und der geringen Anforderungen der Methode, nicht mehr nur auf die Psychologie begrenzt. Beispiele für mögliche Anwendungsbereiche sind Soziologie, Medizin und Trainingsentwicklung für Leistungssportler (Geider, 1982). Generell kann davon ausgegangen werden, dass sich jeder Bereich in dem viele Daten in ein überschaubares und interpretierbares Format gebracht werden sollen der Faktorenanalyse bedienen könnte. Ein großer Vorteil des Verfahrens ist, dass die Faktorenanalyse „als variablenorientiertes (R-Analyse) und als objektorientiertes (Q-Analyses) Datenanalyseverfahren eingesetzt werden“ kann (Bacher, 1996).

Nicht nur der größer gewordene Anwendungsbereich, sondern auch die stetig steigende Rechnerleistung und die zunehmende Benutzerfreundlichkeit von Statistikprogrammen, die immer mehr Forschern die Durchführung dieser Analyse ermöglichen haben zur schnellen Verbreitung der Faktorenanalyse beigetragen.

Der Beginn des Informationszeitalters, die Einführung des elektronischen Zahlungsverkehrs und nicht zu vergessen das Internet stellen Datenmengen von bisher nie da gewesener und unüberschaubarer Größe zur Verfügung. Diese Entwicklungen führen zum Schluss, dass die Bedeutung der datenreduzierenden Analysemethoden noch zunehmen wird.

Nachdem nun die Bedeutsamkeit der Faktorenanalyse hinreichend dargelegt wurde muss allerdings auch auf ihre Risiken hingewiesen werden. Es handelt sich um ein äußerst komplexes Mittel der Analyse und ist nicht innerhalb von wenigen Stunden zu erlernen. Nur durch eingehendes Studium der Fachliteratur und vielfache praktische Anwendung kann das nötige Wissen und die Erfahrung für den Umgang mit dieser Methode erworben werden. Fehlinterpretationen der objektiven Ergebnisse können zu verheerende Folgen führen.

Aus diesem Grunde soll im Anschluss ein grundlegendes Verständnis der faktoranalytischen Denkweise vermittelt werden.

## 2. Explorative Faktorenanalyse

Unter dem Begriff explorative Faktorenanalyse sind eine Menge von Verfahren zusammengefasst, die erschöpfend nur mit einem gehörigen Übermaß an Hybris im Rahmen einer Seminararbeit abgehandelt werden können. Deshalb ruht der Schwerpunkt dieses Abschnitts nicht auf der Erklärung aller Feinheiten, sondern eher auf dem Verstehen des „generellen“ Procedere, als auch auf einigen dabei auftretenden Problemen bzw. Entscheidungsfragen. Das gesetzte Ziel ist ein verantwortungsvoller Umgang mit den mächtigen Verfahren, der eine Basis zur spezielleren

Vertiefung bietet. Bei Bedarf der genauen Rechenvorschriften bzw. der praktischen Umsetzung in Rechenprogrammen sei auf die umfangreichen Werke von Pawlik (1976) und Holm (1976) verwiesen.

Die Arbeitsschritte ergeben sich logisch aus dem in Überla (1971) definierten Hauptziel der Faktorenanalyse. Es ist „die Ableitung hypothetischer Größen oder Faktoren aus einer Menge beobachteter Variablen. Die Faktoren sollen möglichst einfach sein und die Beobachtungen hinreichend genau beschreiben und erklären.“ „Entscheidend ist, dass die Faktorenanalyse eine differenzierte Hypothese über die

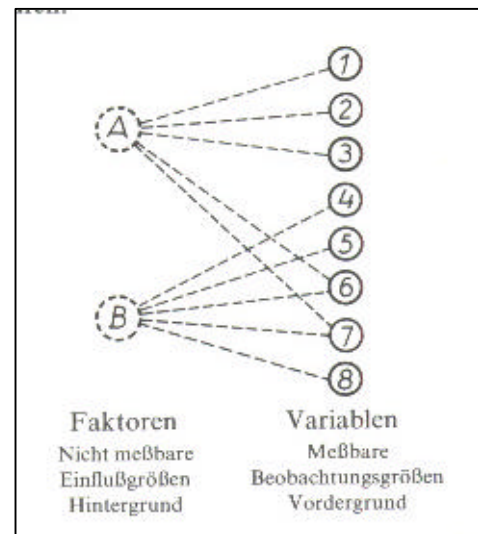


Abbildung 1: Faktorenmodell

Struktur des Zueinander der Variablen möglich macht, ohne dass man vorher eine bestimmte Struktur annehmen oder bereits kennen muss“ (Überla, 1971) (Abbildung 1)

## 2.1 Vorgehen der Hauptkomponentenmethode

Nachdem die Untersuchungsvariablen ausgewählt wurden geht die Hauptkomponentenmethode nach folgendem Schema vor:

1. Bilden einer Korrelationsmatrix
2. Durchführung einer Eigenwertzerlegung:  $R=V \cdot D \cdot V^T$
3. Auswahl der Faktorenzahl
4. Berechnung der Faktorladungen
5. Rotation der Faktoren
6. Für inhaltlich interpretierbare Lösungen können Faktorenwerte für jede Person berechnet werden

Die Aufzählung der einzelnen Schritte wurde sinngemäß aus Bacher (1999) entnommen

## 2.2 Kommunalitäten

Die Kommunalitäten<sup>1</sup> werden im einfachsten Fall als Summe der quadrierten Faktorladungen einer Variable definiert (Bortz, 1999). Ihre Ermittlung fällt unter Punkt 2. in Abschnitt 2.1. Im englischen

<sup>1</sup>  $h_i^2 = a_{i1}^2 + \dots + a_{ir}^2$

Sprachraum finden die Begriffe „Communality“ oder „sum of squared loadings (SSL)“ (Tabachnick & Fidell 1996) Verwendung.

„Die Kommunalität einer Variablen  $i$  gibt an, in welchem Ausmaß die Varianz dieser Variablen durch die Faktoren aufgeklärt wird.“ (Bortz, 1999) Aufgrund der getroffenen Modellannahmen kann die Kommunalität einer Variablen nicht größer als 1 werden. Rein theoretisch wäre es möglich die Faktorenzahl soweit zu erhöhen, dass die komplette Varianz der einzelnen Variablen erklärt werden könnte. Praktisch ist dieses Vorgehen jedoch nicht üblich, da die Variablen meist durch eine geringe Anzahl von Faktoren, bis auf unbedeutende Varianzanteile, erfasst sind. Folglich werden die Kommunalitäten in der Regel kleiner als eins sein (Bortz, 1999).

Tabachnick & Fidell hingegen machen 1996 auf Folgendes aufmerksam: „If communality values equal or exceed 1, problems with the solution are indicated.“ Überla berichtet 199x vom Kommunalitätenproblem. „Die Notwendigkeit, Kommunalitäten wählen zu müssen, die kleiner als 1 sind, d.h. die Existenz des Kommunalitätenproblems, ergibt sich aus dem faktorenanalytischen Modell. Aufbauend auf diesen Überlegungen muss zu dem Schluss gelangt werden, dass die Definition der Kommunalität als Summe der Quadrate der gemeinsamen Faktorenladungen formal und für die Lösung nicht genügend eindeutig ist (Überla, 199x). Auch Pawlik (199x) weist dringlichst auf diese Problematik hin. Ursachen sieht er in der Überbestimmung der Kommunalitäten durch die Verletzung der Annahme, dass die Merkmalsinterkorrelationen unabhängig seien aber auch durch einen, bisher nicht bestimmten Stichrobenfehler. „Man benutzt daher eine der folgenden zwei Näherungsmethoden zur Lösung des Kommunalitäten- und Mindestrangproblems<sup>2</sup>:

1. Kommunalitäten- Schätzmethoden: Die Kommunalitätenmatrix  $\mathbf{H}^2$  wird direkt geschätzt und  $k$  anschließend bestimmt.
2. Mindestrang-Schätzmethoden: Zuerst wird die Zahl  $k$  der gemeinsamen Faktoren abgeschätzt und danach  $\mathbf{H}^2$  berechnet.“ (Pawlik, 1976)

Um die beste Schätzmethode ausmachen zu können ist auf die Zahl der Variablen zu achten. Für kleine Merkmalsstichproben schlägt Pawlik (1976) ein möglichst präzises Verfahren vor, während für große Merkmalsstichproben oft auch gröbere Verfahren ausreichend sind.

## 2.3 Faktorenextraktion

Die Extraktion der Faktoren bzw. Hauptkomponenten beschränkt sich nicht auf ein Verfahren, sondern es steht eine Reihe von mehr oder weniger populären Methoden zur Verfügung. Im Folgenden wird ein kurzer Überblick über die bestehenden Verfahren geliefert. Der zweite Teil dieses Abschnitts wird eine grundsätzliche Hilfe beim Verständnis des Verfahrens liefern.

### 2.3.1 Überblick

In der vorhandenen Literatur wird teilweise, überwiegend im englischen Sprachraum, strikt zwischen principal component analysis (PCA) und factor analysis (FA) unterschieden.

„The two techniques of zwischen principal component analysis (PCA) and factor analysis (FA) are often confuse, particularly in the behavioural sciences, and it is perhaps appropriate to make some comparisons at this stage.“ (Seber, 1984)

„PCA analyses variance; FA analyses covariance (communality). The goal of PCA is to extract maximum variance form a data set with a few orthogonal components. The goal of FA is to reproduce the correlation matrix with a few orthogonal factors. PCA is a unique mathematical solution, whereas most forms of FA are not unique.“(Tabachnik & Fidell 1999)

Weiterhin wird bei Tabachnik & Fidell (1999) auf die verschiedenen Zielsetzungen der Verfahren hingewiesen. PCA wird als das Verfahren, das eine empirische Zusammenfassung des Datensatzes liefert gegen die FA, die eine, von einmaligen and fehlerhaften Variabilitäten „unverseuchte“, theoretische Lösung bietet gestellt.

Gleichwohl werden allerdings auch Lösungen publiziert, die alle Verfahren gleich behandeln (vgl. Pawlik, 1976) oder schlicht die FA erwähnen aber nicht näher erklären Beispielhaft seien (Bortz, 1999; Bacher, 1999). Wir nehmen an, dass die mathematische Eindeutigkeit der Hauptkomponentenanalyse als auch die Tatsache, dass die FA ein lineares Modell, das auf die komplette Korrelationsstruktur Geltung findet, *annimmt*, die Autoren dazu bewegt nur das Verfahren der Hauptkomponentenanalyse genauer zu beschreiben.

Hier soll eine Listung weiterer Verfahren vorgenommen werden, die keinesfalls auf Vollständigkeit besteht:

- principal components
- principal Factors
- Image Factor Extraction
- Maximum Likelihood Factor Extraction
- Unweighted Least Squares Factoring
- Generalized (Weighted) Leas Squares Factoring
- Alpha Factoring

Entnommen aus Tabachnik & Fidell, 1999.

- Hauptkomponentenmethode
- Zentroidmethode
- Algebraische Lösung des Faktorenproblems

---

<sup>2</sup> Zusätzliche ältere Methoden, die heute nicht mehr verwendet werden und daher nicht angeführt sidn, finden sich bei Cattell (1952) und Thorstone (1947)

- Uni-Faktor-Methode
- Two-Faktor-Methode (Spearman)
- Bi-Faktor-Methode
- Multiple Gruppenmethode
- Maximum-Likelihood-Schätzung der Faktorenladungen
- Kanonische Faktorenanalyse
- Alpha-Faktorenanalyse

Entnommen aus Überla, 1999.

In der Diskussion welches Verfahren nun das „richtige“ sei, kommt Überla (1999) zu dem Schluss, dass es in erster Linie auf den „Ausbildungs- und Erfahrungsstand des Untersuchers“ ankommt. Die Problematik, die nötige Rechnerkapazität könnte nicht vorhanden sein, sollte in der heutigen Zeit keine mehr sein. Er schlägt deshalb die *Hauptachsenmethode als Routineverfahren* (Hervorhebung im Original) vor, die sich bei der Extraktion der Faktoren der Hauptkomponentenanalyse bedient. Um ein Gefühl für die Methode zu bekommen empfiehlt er weiterhin das manuelle Rechnen der Zentroidmethode. Zur Wahl der Methode kann mit den Worten von Tabachnik & Fidell (1999) ein guter Abschluss gefunden werden:

„In fact, one test of the stability of a FA solution is that it appears regardless of which extraction technique is employed.“

### 2.3.2 Grundlegende Konzepte

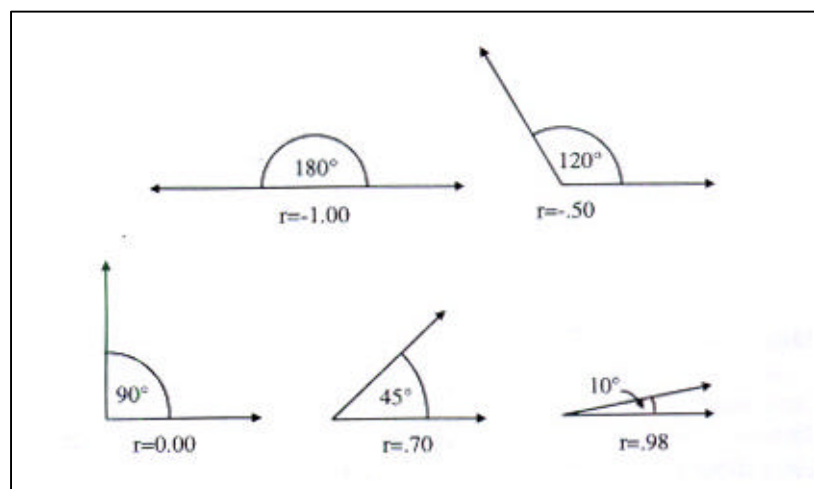
Im Folgenden werden einige Überlegungen dargestellt, die helfen sollen dieses „Gefühl“ für das Verfahren zu bekommen. Die Faktorenanalyse bedient sich der Korrelationen zwischen Variablen.

Abbildung x zeigt die vektorielle Veranschaulichung für verschiedenen Korrelationen. (Abbildung 2)

Bei mehr als drei Dimensionen stößt man natürlich an die Grenzen der geometrischen Darstellbarkeit. Mit Hilfe der abstrakten Matrixform, der sich die Faktorenanalyse bedient,

können diese Fälle jedoch problemlos behandelt werden.

Wir wollen hier allerdings im 3-dimensionalen bleiben, um eine Veranschaulichung zu ermöglichen. In



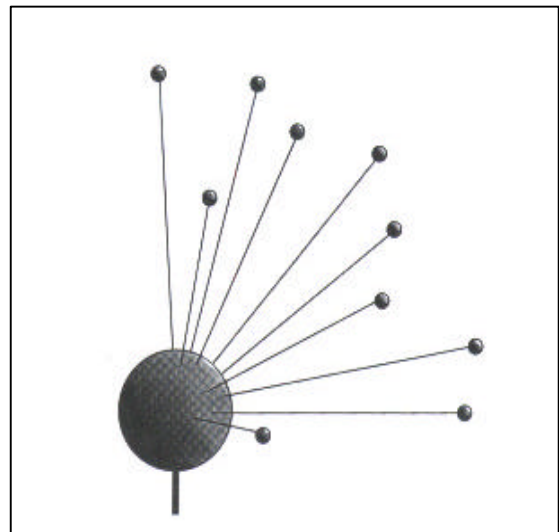
**Abbildung 2:** Veranschaulichung verschiedener Korrelationen

Borg & Staufenbiel (1997) wird das Prinzip der Hauptkomponentenanalyse anhand der Vorstellung von einem Apfel, der mit 10 Stricknadeln zu versehen ist, erläutert. Gegeben ist eine Korrelationsmatrix von zehn Variablen (Tabelle 1):

Aus diesen Korrelationen werden die zugehörigen Winkel errechnet. Jede Variable soll als Stricknadel gesehen werden. Der Apfel stellt den Nullpunkt des Koordinatensystems dar. Nach langem hin- und herprobieren wird man feststellen, dass sich die Stricknadeln nicht nur so anordnen lassen, dass sich die Winkel, die aus der Matrix vorgegeben wurden genau repräsentieren lassen (Abbildung 3), sondern dass sich auch eine 3-dimensionale Vektorkonfiguration (Abbildung 4) finden lässt.

	1	2	3	4	5	6	7	8	9	10
1		16.3	53.1	61.3	50.2	36.9	68.9	61.3	90.0	45.6
2	0.96		36.9	68.9	61.3	53.1	61.3	50.2	90.0	45.6
3	0.60	0.80		90.0	90.0	90.0	53.1	36.9	90.0	60.0
4	0.48	0.36	0.00		16.3	53.1	50.2	61.3	36.9	30.7
5	0.64	0.48	0.00	0.96		36.9	61.3	68.9	53.1	34.9
6	0.80	0.60	0.00	0.60	0.80		90.0	90.0	90.0	60.0
7	0.36	0.48	0.60	0.64	0.48	0.00		16.3	36.9	30.7
8	0.48	0.64	0.80	0.48	0.36	0.00	0.96		53.1	34.9
9	0.00	0.00	0.00	0.80	0.60	0.00	0.80	0.60		45.6
10	0.70	0.70	0.50	0.86	0.82	0.50	0.86	0.82	0.70	

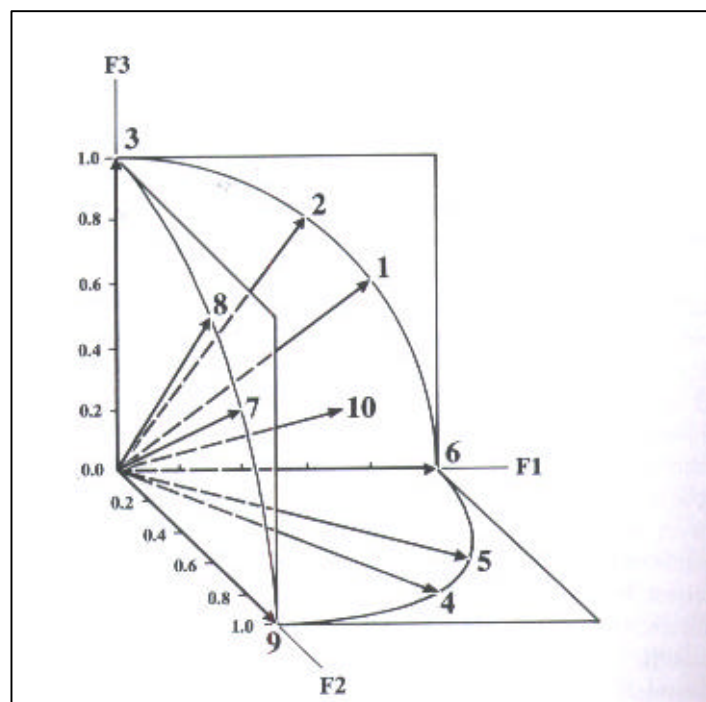
**Tabelle 1:** Korrelationsmatrix



**Abbildung 3:** Stricknadelmodell

### 2.3.3 Anzahl der Faktoren

Weiter oben wurde von einer 3-dimensionalen Struktur geschrieben. Bei der Abhandlung des Kommunalitätenproblems wurde angedeutet, dass durch eine hohe Zahl an Faktoren die Kommunalitäten steigen. Als Ziel der Faktorenanalyse wurde jedoch die Datenreduktion formuliert. Eindeutig handelt es sich bei der Wahl der Faktorenzahl um ein Problem, dass konträre Kriterien in die Entscheidungsfindung einbeziehen muss. Die höchste erklärte Varianz durch die Faktoren wird erreicht durch eine möglichst hohe Zahl



**Abbildung 4:** Vektorenkonfiguration



von Faktoren. In den meisten Fällen lässt sich jedoch die Gesamtvarianz „hinreichend gut“ durch eine Faktorenzahl erklären, die erheblich kleiner ist als die Zahl der Variablen (Bortz, 1999). Es stehen drei wichtige Ansätze zur Festlegung der Faktorenzahl zu Verfügung.

- Kaiser-Guttman-Kriterium

Die Zahl der Faktorenextraktion wird genau dann abgebrochen, wenn der Eigenwert des Faktors, also die erklärte Varianz, unter 1 fällt. Hat ein Faktor einen Eigenwert kleiner 1 würde er weniger erklären als eine die zugrundeliegende Variable. Somit wird der letzte Faktor mit einem Eigenwert größer 1 in die Faktorenlösung einbezogen.

Dieses Verfahren zeichnet sich vor allem durch seine Objektivität aus. Allerdings überschätzt das „KG“-Kriterium oft die Zahl der relevanten Faktoren, was zu Interpretationsschwierigkeiten führen kann

- Scree-Test

Der Scree-Test oder auch das Eigenwertdiagramm liefert weitere Informationen über die Anzahl der zu extrahierenden Faktoren. Die Eigenwerte werden auf der Ordinate, die Abszisse wird mit den Rangnummern belegt. Es bleibt dem Forscher überlassen welchen Eigenwertabfall er als Abbruchkriterium festlegt.

- Parallelanalyse

„Horn (1965) schlägt vor, den Eigenwertverlauf der empirisch ermittelten Korrelationsmatix mit dem Eigenwertverlauf der Korrelation zwischen normalverteilten Zufallsvariablen zu vergleichen (Parallelanalyse).

Entnommen aus Bortz, 1999.

## 2.4 Rotation

Der letzte Schritt vor der Interpretation der Ergebnisse ist die Rotation der Faktorenachsen. Eine unrotierte Faktorenmatrix besitzt so gut wie keinen erklärenden Wert. „Die Rotation wird möglich und nötig, weil die Faktorenextraktion kein eindeutiges Ergebnis erbringt, sondern unendlich viele äquivalente Lösungen, die alle der Gl.  $\mathbf{R}_h = \mathbf{A}\mathbf{A}'$  gleich gut genügen.“ (Überla, 1999). Die Aufgabe der Rotation ist folglich das „richtige“ Koordinatensystem für die Faktorenlösung zu finden. Es finden hierfür nur sog. nicht singuläre Transformationen Verwendung. „Solche Transformationen sind dimensionstreu; sie lassen die Dimension des Raumes unverändert, ...“ (Pawlik, 1976).

Es handelt sich demnach um eine wichtige und folgenreiche Entscheidung welches Rotationsverfahren gewählt wird. In der Forschungspraxis scheinen sich aus verschiedenen Gründen zwei Verfahren etabliert zu haben. Es handelt sich um die Techniken Varimax und Oblimin, die beide in den Bereich der analytischen Rotation fallen (Bortz, 1999). Bei Varimax werden nur orthogonale

Faktoren zugelassen, Oblimin hingegen erlaubt die oblique Rotation. Mit der Begründung, dass „eine entscheidende Funktion der Faktorenanalyse, die Datenreduktion, wieder aufgegeben wird“ (Bortz, 1999) fällt die Entscheidung oft auf orthogonale Rotation. Hinzu kommt, dass die Einbindung eines linear unabhängigen Faktors in ein Geflecht aus Hypothesen leichter zu realisieren ist.

„It should be stressed that factors do not necessarily correlate when an oblique rotation is used.“ (Tabachnik & Fidell, 1999) Holm (1998) geht sogar so weit, dass er publiziert: „Wenn keine besonderen Gründe [s.o.] gegeben sind, dann sollte man schiefwinkelig rotieren.“ Allerdings erfordert die schiefwinkelige Rotation die Eingabe eines Delta- bzw. Gammawertes; je nach verwendetem Programm. Dieser Wert determiniert wie stark die Faktoren untereinander korrelieren dürfen. Wird Delta nicht publiziert und wurde auf  $-4$  eingestellt, so wird sehr wahrscheinlich mit Staunen festgestellt werden, dass trotz obliquer Rotation eine orthogonale Faktorenstruktur gefunden wurde. Durch die Eingabe eines solchen Wertes geht in gewisser Weise ein Stück Objektivität verloren. Ist die Nachvollziehbarkeit durch die korrekte Angabe von Delta bzw. Gamma gewährleistet sprechen nach unserer Einschätzung wenige Gründe gegen die Verwendung von obliquen Rotationstechniken. Wenn manifeste Variable korrelieren können, wäre ein logischer Bruch darin zu finden, dass latente dies nicht vermögen.

## 2.5 Faktorenwerte

Die Faktorenwerte werden wie in einer Regression vergeben. Es ist möglich jeder Vpn einen Faktorenwert zuzuordnen. Nach Überla (1977) wird die Faktorenanalyse oft vor der Berechnung der Faktorenwerte abgebrochen, da ein hoher Rechenaufwand entsteht. Dies ist heute jedoch kein Kriterium mehr, da die Rechnerleistung jedes PCs für die Berechnung der Faktorenwerte ausreicht. Es handelt sich um z-standardisierte Werte auf den extrahierten Faktoren. „The researcher believes that each subject has the same latent factor structure, but different scores on the factors themselves.“ (Tabachnik & Fidell, 1999). Geometrisch ist die Bedeutung der Faktorenwerte, wie sie in Bortz (1999) sehr anschaulich beschrieben wird, folgendermaßen zu verstehen: „Die z-standardisierten Y-Achsen bezeichnen wir als Faktoren und die Koordinaten der Vpn auf den standardisierten Achsen als Faktorenwerte.“ Die Faktorenwerte eines Faktors haben demnach einen Mittelwert von 0 und eine Streuung von 1. Dieses Erkenntnis wird bei der Faktorenschätzung der konfirmatorischen Faktorenanalyse (s.u.) von Bedeutung sein.

## 3 Anwendung des Verfahrens

Als Anwendungsbeispiel für die oben beschriebenen Verfahren wurden Symlog-Daten gewählt. Es handelt sich um ein systematisches, mehrstufiges Gruppenbeobachtungsverfahren (**S**ystematic **m**ultilevel **o**bservation of **g**roups), das von Bales einem Schüler Parsons in jahrelanger Arbeit entwickelt wurde. Nach Bales sind für die Beobachtung von Gruppen 3 Dimensionen

ausschlaggebend. Diese spannen einen Beobachtungsraum auf, der im Rating-Bogen durch eine Skala mit fünf Abstufungen vektorisiert wird.

Für unsere Analysen liegen Datensätze von im Verfahren geschulten und ungeschulten Beobachtern vor. Diese beobachteten in verschiedenen Situationen verschiedene Personen. Hiermit sei klar dargelegt, dass die Daten von einer Unzahl einmaliger und systematischer und unsystematischer Fehlerquellen verwässert sind.

Da zu wenig verschiedene Personen von denselben Beobachtern ratifiziert wurden, beschränkt sich in unserem Fall die Differenzierung der Daten auf geschulte und ungeschulte Beobachter. Ist das der Fall wird auf die Analyse von Teildatensätzen explizit hingewiesen. Werden keine zusätzlichen Angaben zur Stichprobe gegeben, handelt es sich um den kompletten Datensatz aller Beobachter, aller beobachteten Personen und aller Beobachtungssituationen ( $n=218$ ). Die Berechnungen wurden mit ALMO und SPSS durchgeführt. Die gezeigten Tabellen sind mit SPSS erstellt, da die Übertragung von SPSS in Word bei den Autoren bereits eine gewisse Vertrautheit genießt.

Zur Validierung seiner Theorie werden bei Bales (1999) bereits faktoranalytische Ergebnisse veröffentlicht. Nach seinem Willen soll die Grundidee, die hinter seinem Verfahren steckt weiterentwickelt werden. Ausdrücklich wird auf die Wandelbarkeit und Anpassungsmöglichkeiten Symlogs hingewiesen. Den Autoren ist dieser Gedanke völlig verständlich, denn Gruppen setzen sich immer innerhalb eines kulturellen Sinns zusammen. So wird der Raum von Sympathie, Einfluss, und Emotionalität in einer anderen Kultur zwar nicht notwendig aber optional durch unterschiedliche Indikatoren aufgespannt. Bales widmet sich vorwiegend den Polen der Dimensionen, nicht den Dimensionen selbst. Der Definition dieser Richtungskodierung werden 56 Seiten (ca. 8% des Gesamtwerkes) beigemessen. Die Faktorenanalyse soll hier eingesetzt werden, um reflektierend zu einer Einschätzung des Modells zu gelangen. Im einem folgenden Schritt können daraus eventuell Verbesserung des Beobachtungsverfahrens abgeleitet werden.

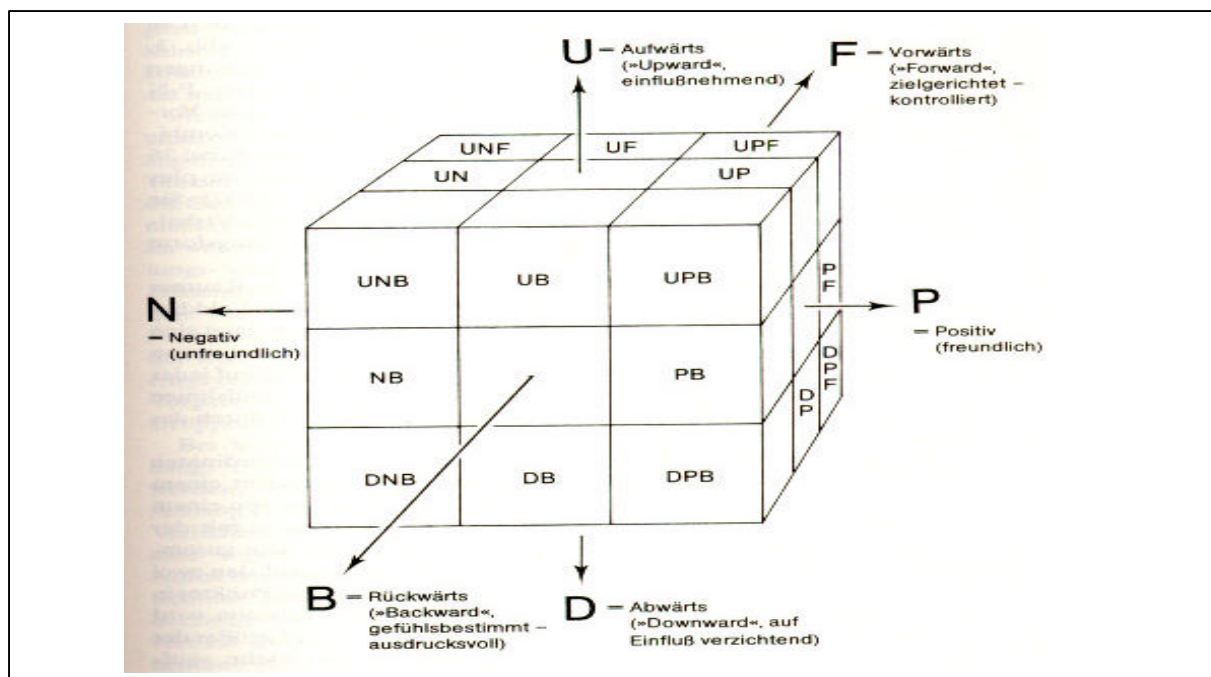


Abbildung 5: Der Symlograum

### 3.1 Analyse der dimensionsbezeichnenden Items

Um zu möglichst klaren Ergebnissen zu gelangen wurden für die ersten Analysen nur die Variablen verwandt, die theoretisch direkt auf den zu extrahierenden Dimensionen liegen sollen (Abbildung 5). Es handelt sich hierbei um:

U: aktiv, dominant, spricht viel

P: freundlich, partnerschaftlich

F: analytisch, aufgabenorientiert, lösungsorientiert

N: unfreundlich, negativistisch, individualistisch

B: emotional, spontan

D: passiv, introvertiert, spricht wenig

Für diese Ausgangskonfiguration wurden 14 Verfahren gerechnet. Zur Extraktion der Faktoren wurde Hauptkomponentenanalyse, nicht gewichtete kleinste Quadrate, verallgemeinerte kleinste Quadrate, Maximum-Likelihood, Hauptachsen-Faktorenanalyse, Alpha-Faktorisierung und Image-Faktorisierung gerechnet. Rotiert wurde jeweils mit Varimax und Oblimin (Delta=0).

Die Alpha-Faktorisierung ergab keine Lösung, da nach Iteration 25 eine Kommunalität den Wert 1 überschritt. Die Extraktion wurde abgebrochen. Die verbleibenden 12 Ergebnisse erbrachten alle mehr oder weniger eindeutig die 3 von Bales (1982) beschriebenen Dimensionen. Die Variable F wurde von der Image-Faktorisierung nicht eindeutig zugeordnet. Sonst wurden die Pole in jeder Lösung korrekt in einen gemeinsamen Faktor gefasst. Hierbei war wiederum eine Variabilität der erklärten Varianz festzustellen. Beispielhaft zeigen Abbildung xx und xx die beste und schlechteste Extraktion des Modells (Varimax-Rotation). Das Kaiser Kriterium erbrachte eindeutig eine 3 Faktorenlösung.

**Rotierte Komponentenmatrix**

	Komponente		
	1	2	3
U	-,314	,765	5,156E-02
P	,862	-,206	,206
F	,469	,477	-,418
N	-,870	6,381E-02	6,165E-02
B	,101	8,667E-02	,942
D	3,511E-03	-,764	-2,86E-02

Extraktionsmethode: Hauptkomponentenanalyse.

Rotationsmethode: Varimax mit Kaiser-Normalisierung.

a. Die Rotation ist in 6 Iterationen konvergiert.

**Tabelle 2:** Hauptkomponentenanalyse

Mittels Varimax-Rotation und Hauptkomponentenanalyse (Tabelle 2) konnten über 70% der Gesamtvarianz erklärt werden. P nimmt Werte um .8 an und N -.8. Für U werden Werte von .7 ermittelt, D erhält geringere Werte. D schwankt zwischen -.319 und -.764. In den meisten Fällen erhält D ca. -.5. Die Dimension FB erbringt die schlechtesten Ergebnisse. Weder sind die Werte in etwa

gleich groß, noch sind sie eindimensional. B legt mit Werten von bis zu .999 die Dimension fest. Bis auf die Image-Faktorisierung (Tabelle 3) und (0.21) und die Hauptachsen-Faktorenanalyse (.651) erhält B immer Ladungen über .94. F lädt zwar immer negativ auf den entsprechenden Faktoren, doch sind auf den anderen zwei Faktoren von positiven F Werten mit höherer Ausprägung zu berichten.

Die verschiedenen Rotationen zeigten allerdings Unterschiede in der „Wichtigkeit“ der Faktoren bzw. Komponenten. Die Dimension PN wurde in allen 6 Ergebnissen des Varimax-Rotationsverfahrens als Faktor mit der höchsten aufgeklärten Varianz ermittelt. FB bzw. UD wurden je dreimal als zweiter und dreimal als dritter Faktor ermittelt.

**Rotierte Faktorenmatrix<sup>a</sup>**

	Faktor		
	1	2	3
U	-,169	,408	-6,28E-02
P	,553	-,304	,189
F	,272	,126	-6,07E-02
N	-,579	,218	-9,70E-02
B	1,615E-02	-3,17E-02	,219
D	-1,98E-02	-,319	4,362E-02

Extraktionsmethode: Image-Faktorisierung.

Rotationsmethode: Varimax mit Kaiser-Normalisierung.

a. Die Rotation ist in 5 Iterationen konvergiert.

**Tabelle 3:** Image-Faktorisierung

Die mittels Oblimin erhaltenen Lösungen erbrachten folgende Konfiguration.

Faktor	PN	FB	UD
1	4	2	0
2	2	1	3
3	0	3	3

**Tabelle 4:** Oblimin-Lösungen

Somit sehen die Autoren es als eindeutig an, dass innerhalb des verwandten Datensatzes die Sympathie (PN) die höchste Erklärungskraft der Varianz aufweist. Diese wird von einer nicht immer eindeutig bestimmbar Zielgerichtetheit (FB) gefolgt. Einfluss (UD), dem bei Bales ein sehr hoher Stellenwert zukommt, wird hier empirisch als die Dimension mit der geringsten Erklärungskraft ermittelt.

Die Analyse der Korrelationsmatrix (Tabelle 5) der Dimensionen kommt ebenfalls zu interpretierbaren und übereinstimmenden Ergebnissen. Bis auf die Image-Faktorisierung, die bereits bei der Extraktion von FB Ergebnissen kam, die nicht mit den übrigen Verfahren übereinstimmten, wurde stets eine negative Korrelation von UD und FB mit der Stärke knapp unter  $r=0.2$  ermittelt. Die übrigen Korrelationen sind derart gering, dass lineare Unabhängigkeit vermutet werden kann.

Dennoch stellen wir fest, dass F wie Oben beschrieben und die Variable U ebenfalls auf dem Faktor (PN) mit ca. -.22 lädt. Aus diesem Grund wurde erneut eine Hauptkomponentenanalyse mit Oblimin-Rotation gerechnet. Um eine erhöhte „Korrelationsfreude“ zu erreichen wurde Delta auf 0.6 festgelegt.

**Komponentenkorrelationsmatrix**

Komponente	1	2	3
1	1,000	7,840E-02	-,381
2	7,840E-02	1,000	-,463
3	-,381	-,463	1,000

Extraktionsmethode: Hauptkomponentenanalyse.

Rotationsmethode: Oblimin mit Kaiser-Normalisierung.

**Tabelle 5:** Komponentenkorrelationsmatrix

Es kann gezeigt werden, dass Komponente drei, der wie Oben beschrieben die Dimension (FB) darstellt nicht senkrecht auf den Komponenten 1 und 2 steht, während diese untereinander nahezu linear unabhängig sind. Wird Delta erneut um 0.1 erhöht sind alle Korrelationen über .86. Mit einem Delta von .8 kommt es zu einem Abbruch des Verfahrens.

Wird nun aus theoretischen Gründen eine Lösung mit linear unabhängigen Faktoren gewünscht, so sollte erneut über die Items F und B reflektiert werden. Bales (1982) begnügt sich in seinem Raum-Modell nicht damit nur die grundlegenden Dimensionen durch Items zu vektorisieren. Auch Mischdimensionen, die sich z.B. aus den Items DP und UN zusammensetzen können werden im Adjektiv-Ratingbogen erhoben.

### 3.2 Erweiterung auf 18 Items

Da die Grunddimensionen aus den Daten zufriedenstellend extrahiert werden konnten, wurde im nächsten Schritt die Menge der zu untersuchenden Variablen um die zweidimensionalen Items erweitert.

UP: extravertiert, geht aus sich heraus, sicher, beliebt;

UF: tatkräftig, durchsetzungsfreudig;

UN: dominant, eigensinnig, nachdrücklich;

UB: macht Späße, schauspielert, geht aus sich heraus;

PF: interessiert, kooperativ;

NF: kritisch, gewissenhaft, prinzipiell;

NB: uninteressiert, unwillig, nicht kooperativ;

PB: warmherzig, natürlich, freundschaftlich;

DP: verständnisvoll, tolerant, gelassen;

DF: besonnen, sachlich;

DN: traurig, niedergeschlagen, deprimiert;

DB: unentschlossen, ängstlich;

Wegen der erhöhten Datenmenge wurden zwar alle oben angewandten Verfahren auch für die 18 Items gerechnet. Wegen der nach oberflächlicher Analyse festgestellten relativen Ähnlichkeit der Ergebnisse wurde sich zur genaueren Analyse auf die Hauptkomponentenanalyse mit Varimax- und Oblimin-Rotation beschränkt.

### 3.2.1 4 Faktorenlösung

Ein grundlegender Unterschied zu den Ergebnissen aus Abschnitt 3.1 ist die Extraktion von vier Komponenten nach dem Kaiser-Kriterium (Tabelle 6). Diese vier Komponenten erklären zusammen 64% der Varianz. Würde man sich auf eine 3 Komponentenlösung festlegen, würde dies zu einem Verlust von ca. 10% der erklärten Varianz führen. Auch Bales (1982) extrahiert diesen „vierten Faktor“. Die erklärte Varianz dieses Faktors gibt er mit 6.4% an. Weiterhin führt er aus, dass dieser Faktor bei geschulten Ratern kaum noch eine Rolle spielt. „Dies ist ein weiterer Hinweis darauf, dass das Training zu einer Umstrukturierung der Itembedeutung führte.“ (Bales, 1982) Er weist auf die schlechte Interpretierbarkeit dieses Faktors hin und bezieht sich auf Wish, der ebenfalls einen solchen Faktor extrahierte und ihn mit „Intensität“ bezeichnet. Wie in unseren Ergebnissen klärt dieser vierte Faktor bei Wish einen größeren Varianzanteil als 6.4% auf.

Die schlechte inhaltliche Interpretation der Komponenten ist auch in unserem Fall zu beklagen.

**Rotierte Komponentenmatrix<sup>a</sup>**

	Komponente			
	1	2	3	4
U	-,413	,480	-,243	,188
P	,818	-,128	,122	,348
F	,383	,727	-5,85E-02	-9,56E-02
N	-,818	,186	,123	-1,54E-02
B	-4,76E-02	-7,83E-02	6,229E-02	,750
D	4,658E-02	-,146	,851	-2,78E-02
UP	,289	,265	-9,11E-02	,651
UF	-,132	,627	-,389	,174
UN	-,703	,493	-,100	,131
UB	,172	-4,19E-03	-6,29E-02	,713
PF	,746	,233	4,578E-02	,213
NF	-,230	,758	-6,11E-02	4,398E-04
NB	-,647	-6,73E-03	,252	8,832E-02
PB	,744	-4,37E-02	,215	,381
DP	,810	2,595E-02	,125	,159
DF	,504	,487	5,900E-02	-,289
DN	-6,21E-02	2,534E-02	,817	7,565E-02
DB	6,759E-02	-,209	,750	-8,04E-02

Extraktionsmethode: Hauptkomponentenanalyse.

Rotationsmethode: Varimax mit Kaiser-Normalisierung.

a. Die Rotation ist in 11 Iterationen konvergiert.

**Tabelle 6:** Rotierte Komponentenmatrix

Die meisten Ladungen sind eindeutig auf Komponente 1 zu verorten (P, N, UN, PF, NB, PB, DP). Jedes auf Item, das auf Komponente 1 hoch lädt beinhaltet die Dimension Sympathie. Dies würde sich mit den Ergebnissen aus 3.1 decken. Auch sind in dieser Komponente die eindimensionalen Items (P, N) der Sympathie. Weitere eindimensionale Items treten nicht auf, auch wenn U eine relativ hohe Ladung zugeordnet wird.

Komponente 2 ( $u^3$ , F, UF, NF, df, un) vereint die Grundrichtungen U und F miteinander. F kommt eindeutig eine höherer Stellenwert zu. Das Item df müsste, wenn nur eine rein statistische Interpretation der Ergebnisse vorgenommen würde, zur ersten Komponente gezählt werden. Inhaltlich passt dieses Item besser zu Komponente zwei.

Komponente 3 (D, DN, DB, u, uf,) kann nun als die dritte Dimension des Symlog-Raumes (Einfluss) gesehen werden. Item uf in dieser Komponente zu nennen ist, aufgrund dessen, dass es eigentlich schon zu Komponente 2 gezählt wurde, fraglich. Hier muss wieder inhaltlich argumentiert werden.

Trotz der auftretenden Schwierigkeiten können doch wenigstens tendenziell die Ergebnisse des vorigen Abschnitts bestätigt werden.

Komponente 4 (B, UP, UB) als Intensität zu bezeichnen wäre, wenn man sich die Adjektive aus dem Ratingbogen erneut betrachtet (emotional, spontan, extravertiert, geht aus sich heraus (2mal), sicher, beliebt, macht Späße, schauspielert), nicht sehr abwegig.

### 3.2.2 3 Faktorenlösung

Dennoch wurde in einer weiteren Untersuchung die Faktorenzahl manuell auf drei festgelegt (Tabelle 7). Somit werden letztendlich nur 55,6% der Varianz erklärt. Es wird die mit Varimax erreichte Komponentenmatrix genauer analysiert, da das zugrundeliegende theoretische Modell von einem orthogonalen Raum ausgeht.

Die Ergebnisse lassen die Faktorenstruktur erneut vermuten.

Komponente 1 lädt auf den Variablen P und N hoch. Weiterhin kann bei allen Items, die ein P beinhalten ein positiver und bei allen, die ein N beinhalten eine negative Ladung festgestellt werden. Im Optimalfall würden die Items P und N mit 1.0 und die zweidimensionalen Items mit 0.5 laden. Dieser ist, auch das ist eindeutig zu erkennen, nicht erreicht. Hier bietet die Faktorenanalyse gute Möglichkeiten die Schwächen des Modells, seien sie durch Messfehler, Verzerrung durch die Stichprobe oder schlechter Operationalisierung entstanden aufzuspüren. In einem zweiten Schritt können die Ergebnisse zur Verbesserung der Modellanpassung führen.

---

<sup>3</sup> Klein geschriebene Items laden weniger stark



**Rotierte Komponentenmatrix<sup>a</sup>**

	Komponente		
	1	2	3
U	-,413	,525	,211
UP	,260	,236	,674
UF	-,124	,719	,221
UN	-,713	,442	,175
UB	,148	3,572E-02	,662
P	,792	-,212	,379
PF	,720	9,447E-02	,321
F	,364	,526	,109
NF	-,250	,575	,163
N	-,831	7,427E-02	5,116E-03
NB	-,670	-,159	9,551E-02
B	-8,12E-02	-9,64E-02	,692
PB	,705	-,218	,448
DP	,786	-,108	,242
DF	,491	,270	-9,66E-02
DN	-,142	-,572	,268
DB	7,825E-03	-,690	5,525E-02
D	-2,69E-02	-,719	,143

Extraktionsmethode: Hauptkomponentenanalyse.

Rotationsmethode: Varimax mit Kaiser-Normalisierung.

a. Die Rotation ist in 6 Iterationen konvergiert.

**Tabelle 7: Rotierte Komponentenmatrix**

Komponente 2 bildet wieder UD ab. Bis auf Item DF zieht stimmen die Vorzeichen aller zweidimensionalen Items in Komponente 2 mit denen der eindimensionalen Items überein. Die Ladungen weichen allerdings wieder von den theoretisch geforderten ab.

Komponente 3 beinhaltet bis auf das, auch theoretisch negativ vermutete, Item DF, welches sehr schwach lädt, nur positive Ladungen und widerspricht somit schon in diesem Kriterium den theoretisch erwarteten Werten. Auch die Behauptung die Items, die ein F beinhalten würden tendenziell niedriger laden als solche deren Name ein B beinhaltet lässt sich nicht halten. Aber auch dieses Ergebnis stimmt mit den in Abschnitt 3.1 ermittelten Ergebnissen überein.

### 3.3 Analyse des gesamten Symlog-Raums (alle 26 Items)

Wir haben gesehen, dass durch die Zuführung weiterer Variabler die Schärfe der Ergebnisse nachlässt. Kaum verwunderlich ist es daher, dass die in der Komplettkonfiguration – alle 26 Symlog-Items – Ergebnisse noch schwieriger zu interpretieren sind. Es kommen zu den bereits gelisteten Items noch

UPF: zielbewusster und aufgabenorientierter, demokratischer Leiter;

UNF: disziplinierend, folgerichtig;

UNB: geltungssuchend, selbstbezogen, provozierend;

UPB: optimistisch, humorvoll, hilfsbereit;

DPF: rücksichtnehmend, zuverlässig, andere anerkennend;

DNF: selbstkritisch, pflichtbewusst;

DNB: entmutigt, verletzt, resignierend;

DPB: behaglich, gemütlich, zufrieden;

hinzu.

Nach dem Kriterium, dass der Eigenwert eines Faktors größer eins sein muss, erhalten wir eine fünffaktorielle Lösung (Tabelle 8):

**Rotierte Komponentenmatrix<sup>a</sup>**

	Komponente				
	1	2	3	4	5
U	-,212	,615	-6,33E-02	,176	-,311
UP	,401	,311	2,540E-02	,627	-,146
UPF	,739	,341	-,100	6,253E-02	-,137
UF	6,588E-02	,708	-,268	,129	-,183
UNF	-,327	,661	-,138	-,131	9,806E-02
UN	-,609	,645	-5,53E-02	8,098E-02	,101
UNB	-,693	,442	,117	,153	-3,96E-02
UB	,171	5,448E-03	-5,38E-02	,720	3,575E-02
UPB	,725	-,185	7,157E-02	,360	9,611E-02
P	,751	-,282	5,229E-02	,344	,182
PF	,758	8,985E-02	3,557E-02	,187	,104
F	,521	,551	-9,11E-02	-,130	,134
NF	-9,94E-02	,730	-9,76E-02	-,103	,259
N	-,787	,306	9,842E-02	-1,31E-02	9,458E-02
NB	-,641	7,987E-02	,223	,121	,175
B	-,107	-3,95E-02	-3,86E-02	,680	,309
PB	,682	-,210	,113	,351	,392
DP	,738	-,173	2,269E-02	,141	,374
DPF	,767	-,160	9,369E-02	6,579E-02	,285
DF	,505	,273	-8,17E-03	-,331	,372
DNF	,513	8,433E-02	7,490E-02	-,349	,235
DN	-1,85E-02	-1,97E-02	,862	2,759E-02	2,159E-02
DNB	-,103	3,136E-02	,874	-5,14E-03	-4,26E-02
DB	5,912E-02	-,287	,738	-6,47E-02	-2,39E-02
DPB	,222	7,977E-02	,105	,221	,725
D	-2,83E-03	-,276	,729	-5,19E-02	,243

Extraktionsmethode: Hauptkomponentenanalyse.

Rotationsmethode: Varimax mit Kaiser-Normalisierung.

a. Die Rotation ist in 12 Iterationen konvergiert.

**Tabelle 8:** Fünffaktorenlösung

Die Behauptung, dass es sich bei Komponente 1 um die Dimension PN handelt kann aufgrund der hohen Ladung der eindimensionalen Items P und N und den nahezu konstant eingehaltenen Vorzeichen der übrigen Items untermauert werden.

Komponente 2 als Dimension UD zu bezeichnen fällt nicht nur aufgrund einer NF von .730 schon schwerer. Die negativen Ladungen, die von den Items, die ein D beinhalten erwartet werden müssen, sind unserer Ansicht nach zu gering. Hingegen der erwarteten Ladungen auf F bzw. B Items laden in Komponente 3 D Items überdurchschnittlich hoch. Adhoc könnte von einer Trennung der Dimension UD in zwei eigenständige Komponenten ausgegangen werden.

Mit sehr viel gutem Willen laden in Komponente 4 B Items positiv und F Items negativ. Aber auch P und UP weisen hohe Ladungen auf. Komponente 5 inhaltlich zu verwerfen stellt ein stellt ein den Autoren derzeit unlösbares Problem dar.

Werden bei der Extraktion der Faktoren wieder 3 vorgegeben, so ergeben sich ähnliche Ergebnisse (Die Ladungen werden nicht abgebildet). Die erste Komponente kann relativ sicher als Dimension PN ausgemacht werden. Komponente 2 und 3 lassen sich wieder mit viel Phantasie als U bzw. D verstehen. Hierbei handelt es sich in keinem Fall um eine „schöne“ Faktorenlösung.

### **3.4 Zusammenfassung**

Wir haben grundlegend dargestellt, wie das Verfahren explorative Faktorenanalyse funktioniert. Die Vielfalt der Einsatzmöglichkeiten für diese Methode der Datenanalyse sollte dem Leser bewusst geworden sein. An einem konkreten Datensatz wurden Ergebnisse der Faktorenanalyse gezeigt und inhaltlich interpretiert. Schließlich wurde die Möglichkeit angesprochen die Ergebnisse zur Verbesserung der Operationalisierung einzelner Items heranzuziehen.

Die Interpretation der Ergebnisse ergab eine relativ eindeutig Extraktion der Dimensionen PN und UD. Die Dimension FB mit weniger guten Ergebnissen dennoch gefunden. Jedoch passt auch dieses Ergebnis zu den von Bales (1983) beschriebenen.

Mit der höherer Anzahl der in die Analyse eingegangenen Items wurden die Ergebnisse uneindeutiger, bis letztlich die theoretisch zugrundeliegenden Dimensionen nahezu nicht mehr zu finden sind. Auch die erklärte Varianz nimmt mit steigender Itemzahl ab.

## **4 konfirmatorische Faktorenanalyse**

In den folgenden Abschnitten der vorliegenden Arbeit wird das Verfahren der konfirmatorischen Faktorenanalyse, ein modellorientiertes, hypothesentestendes Verfahren für Strukturgleichungen der Faktorenanalyse vorgestellt und eingesetzt. Im Rahmen der datenorientierten Analyse ist die, bereits beschriebene, exploratorische Faktorenanalyse von größerer Bedeutung (Roth, 1984). Die hier kurz

vorgestellte und anschließend verwendete Software AMOS realisiert strukturelles Modellieren in der Tradition Jöreskog, welcher auch schon das SPSS-Modul LISREL geschaffen hat.<sup>4</sup>

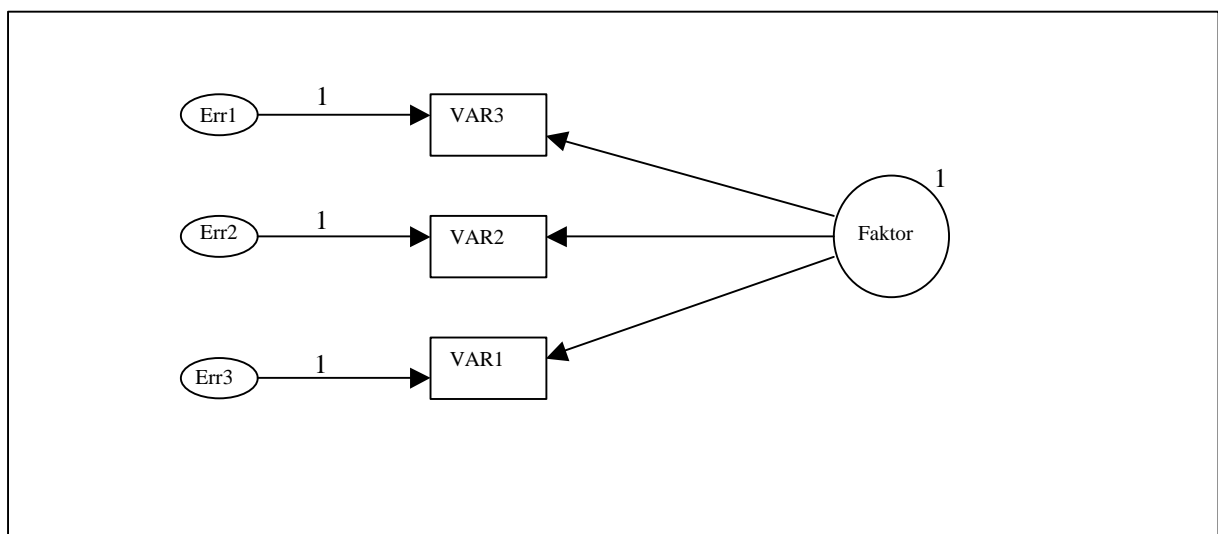
Das Modellieren von Strukturgleichungen hat den Ruf, ein besonders schwieriges Verfahren zu sein, welches nur von einigen Überfliegern oder Super-Statistik-Gurus verstanden und angewendet werden kann. In diese Arbeit wollen wir aber zeigen, daß dies nicht der Fall ist, da AMOS eine äußerst benutzerfreundliche Bedienungsoberfläche besitzt und die Umsetzung der Modelle fast rein intuitiv geschieht. Nichtsdestotrotz ist einiges an statistischem und linear-algebraischem Wissen von Nöten.

## 4.1 Kurze Einführung

AMOS besitzt zwei Arten von Steuerung: Die oben schon erwähnte, benutzerfreundliche, grafische Oberfläche und eine ebenso leicht zu bedienende Steuerung durch sogenannte „dollar sign“ (\$) Kommandos.

### 4.1.1 Grafische Steuerung

Die grafische Oberfläche läßt sich bedienen wie ein konventionelles, einfaches Zeichenprogramm, wie man es beispielsweise bei der Windows-Standard-Installation unter Paint-Brush findet. Dargestellt werden beobachtete (observed; tatsächlich vorhandenen) Variablen als Vierecke und die in unserem Anwendungsfall latenten Faktoren als Kreise oder ovale Gebilde. Nicht gemessene d.h. nicht beobachtete Fehlereinflüsse werden auch als Kreis oder elliptische Gebilde dargestellt. Das Beispiel in Abbildung 6 soll das ein wenig deutlicher machen:

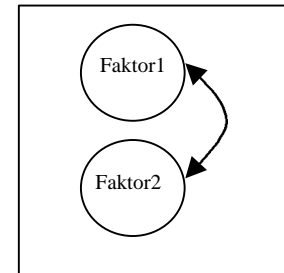


**Abbildung 6:** einfaches Beispiel

<sup>4</sup> Eine Studentenversion, die allerdings auf acht „observed“ Variablen beschränkt ist, steht kostenlos zum download auf <http://www.smallwaters.com/amos> bereit.

Das hier gezeigte Beispiel mit drei Variablen, drei Fehlereinflüsse und einem Faktor soll das Prinzip der Vorgehensweise bei der grafischen Oberfläche bei AMOS verdeutlichen. Das zu testende Modell wird aufgrund theoretischer Überlegungen aufgebaut. In den drei Variablen liegt eine gemeinsame, latente Variable zugrunde, nämlich der Faktor. Da man in der empirischen Forschung aber niemals perfekte Daten erhebt, werden den Variablen noch Fehlereinflüsse hinzugefügt.

Die einköpfigen Pfeile stellen die Regressionskoeffizienten<sup>5</sup> der Variablen auf den Faktor dar (z.B.  $a_1$ ,  $a_2$  und  $a_3$ ). Zweiköpfige, gebogene Pfeile stellen die Korrelationskoeffizienten dar. Im Kasten rechts ist beispielhaft eine Korrelation zwischen Faktor 1 und Faktor 2 abgebildet.



#### 4.1.2 Kommando Steuerung

Die Steuerung durch \$commands ist fast ebenso einfach wie die grafische Steuerung. Das oben dargestellte Modell lässt sich beispielsweise durch folgende Steuerbefehle beschreiben:

\$Structure

```
Variable1 < ----- Faktor1
Variable2 < ----- Faktor1
Variable3 < ----- Faktor1
Variable1 < ----- error1 (1)
Variable2 < ----- error2 (1)
Variable3 < ----- error3 (1)
Faktor1 (1)
```

Nachdem nun ein Modell entworfen wurde, wird es mit der Struktur der erhobenen Daten verglichen. Das ist der eigentliche Test. Hier stehen einem mit AMOS natürlich eine Reihe von Optionen, vor allem was die Ergebnispräsentation anbelangt, zur Auswahl.

Bevor jedoch mit der eigentlichen Berechnung begonnen werden kann, muss das entworfene Modell, welche ja aus einer Reihe von Gleichungen besteht, identifiziert werden. D.h. das Gleichungssystem muss gelöst werden können.

<sup>5</sup> Regressionskoeffizienten deshalb, weil die Berechnungen auf der Grundlage der Regression basieren. Näheres in Kapitel Identifikation.

## 4.2 Identifikation

Die grundlegendste Regel ist die t-Regel. Sie besagt, dass die Zahl der unbekannten Modellparameter kleiner/gleich der Zahl der bekannten Modellparameter sein muss (Bollen, 1989; Entnommen aus Bacher, Vorlesungsskript). Dies ist eine notwendige Bedingung ohne die die Identifikation nicht möglich ist, denn es ist auf eine eindeutige Art auch nicht möglich eine Gleichung mit zwei unbekannten Parametern zu lösen. Damit ist das Modell aber noch nicht identifiziert.

Die eigentliche Identifikation des Modells kann man anhand von zwei weiteren Regeln vornehmen, aber in manchen Fällen muss auch Hand angelegt werden um das Modell zu identifizieren. Dies ist der Fall, wenn die zwei Identifikationsregeln nicht zutreffen, denn es handelt sich bei den Bedingungen der Regeln um hinreichende Bedingungen und nicht um notwendige Bedingungen. In klaren Worten ausgedrückt bedeutet das: wenn die Bedingungen erfüllt sind, so ist das Modell identifiziert. Umgekehrt jedoch, wenn die Bedingungen nicht erfüllt sind, kann über das Modell keine Aussage getroffen werden. Es ist dann die Identifikation im konkreten Fall zu prüfen.

Die zwei Regeln lauten wie folgt:

Identifikationsregel	Bedingungen
3-Indikatorenregel	<ul style="list-style-type: none"> <li>• Es liegen ein oder mehrere Faktoren vor.</li> <li>• Jeder Faktor wird durch mindestens drei Variablen gemessen, d.h., dass mindestens drei Variablen <b>nur</b> auf diesen Faktor laden.</li> <li>• Die Meßfehler sind paarweise untereinander und von den gemeinsamen Faktoren unabhängig</li> </ul>
2-Indikatorenregel	<ul style="list-style-type: none"> <li>• Es liegen mindestens <b>zwei Faktoren</b> vor, die untereinander <b>korreliert</b> sind.</li> <li>• Jeder Faktor wird durch mindestens <b>zwei Variablen</b> gemessen, d.h., dass mindestens zwei Variablen <b>nur</b> auf diesen Faktor laden.</li> <li>• Die Meßfehler sind paarweise untereinander und von den gemeinsamen Faktoren unabhängig</li> </ul>

**Tabelle 9:** Identifikationsregeln

Diese Regeln werden plausibel, wenn man sich die Struktur des Gleichungssystems anschaut. Die Faktorenanalyse wird, wie in der Fußnote schon erwähnt, auf der Basis der linearen Regression gerechnet. Demnach lassen sich die Variablen durch eine lineare Kombination von Regressionskoeffizienten bzw. Faktorladungen, Faktor und Fehlereinfluß darstellen. (siehe Beispiel oben \$commands) Aus dem Beispiel in Abbildung 6 folgt dann:

$$\text{Variable1} = a_1 \times f + \text{error}_1$$

$$\text{Variable2} = a_2 \times f + \text{error}_2$$

$$\text{Variable3} = a_3 \times f + \text{error}_3$$

Auf dieser Basis lassen sich dann die erwarteten Varianzen und Kovarianzen der beobachteten (vorhandenen) Variablen berechnen.

$$\text{VAR}(\text{Variable1}) = a_1^2 \times \text{VAR}(f) + \text{VAR}(\text{error}_1)$$

$$\text{VAR}(\text{Variable2}) = a_2^2 \times \text{VAR}(f) + \text{VAR}(\text{error}_2)$$

$$\text{VAR}(\text{Variable3}) = a_3^2 \times \text{VAR}(f) + \text{VAR}(\text{error}_3)$$

$$\text{KOV}(\text{Variable1}, \text{Variable2}) = a_1 \times a_2 \times \text{VAR}(f)$$

$$\text{KOV}(\text{Variable2}, \text{Variable3}) = a_2 \times a_3 \times \text{VAR}(f)$$

$$\text{KOV}(\text{Variable1}, \text{Variable3}) = a_1 \times a_3 \times \text{VAR}(f)$$

Die fettgedruckten Elemente des Gleichungssystems sind bekannt. Man sieht, dass das System mit diesen Angaben noch nicht zu lösen ist, denn wir haben 7 Unbekannte welche da sind:  $a_1$ ,  $a_2$ ,  $a_3$ ,  $\text{VAR}(f)$ ,  $\text{VAR}(\text{error}_1)$ ,  $\text{VAR}(\text{error}_2)$  und  $\text{VAR}(\text{error}_3)$ .

Um nun das Modell zu identifizieren, können die Varianzen der Faktoren gleich 1 oder bestimmte Faktorladungen gleich 1 gesetzt werden. Das als zweites genannte Vorgehen ist das übliche Vorgehen in der konfirmatorischen Faktorenanalyse. Man erhält dadurch 7 Bekannte und 6 Unbekannte, da nun entweder  $a_1$  oder  $\text{VAR}(f)$  gegeben ist. AMOS ist in der Lage das Modell empirisch zu testen. Gibt es Schwierigkeiten mit Identifikation, so testet AMOS das Modell und meldet, wenn das Modell nicht identifiziert ist. Allerdings ist diese Methode kein Ersatz für die tatsächliche empirische Identifikation des Modells.<sup>6</sup>

Problematisch wird es dann z.B., wenn nun nur zwei Variablen auf einem Faktor laden. so hat man nur die Gleichungen:

$$\text{VAR}(\text{Variable1}) = a_1^2 \times \text{VAR}(f) + \text{VAR}(\text{error}_1)$$

$$\text{VAR}(\text{Variable2}) = a_2^2 \times \text{VAR}(f) + \text{VAR}(\text{error}_2)$$

mit der Kovarianz:

$$\text{KOV}(\text{Variable1}, \text{Variable2}) = a_1 \times a_2 \times \text{VAR}(f)$$

zur Verfügung.

<sup>6</sup> Ein Modell kann zwar theoretisch identifiziert, dennoch empirisch nicht identifiziert sein. Dies ist z.B. in einem 2-Faktorenmodell der Fall, wenn zwar jeweils zwei Variablen einen Faktor messen, die Faktoren aber empirisch unkorreliert sind.

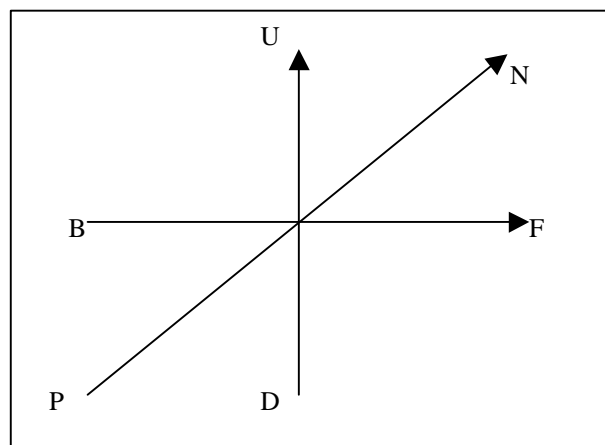
Das Gleichungssystem ist nun nicht mehr lösbar, d.h. nicht identifizierbar. Deshalb wird bei der 2-Indikatorenregel gefordert, dass mindestens 2 Faktoren, die miteinander korrelieren, vorhanden sind. Es können so über den zweiten Faktor Beziehungen zu den Variablen, welche auf dem ersten Faktor laden hergestellt werden.

Zu Bemerken ist hier aber jedoch die Ausnahmestellung der „einfachen“ Modelle. Passen wie gesagt die Identifikationsregeln nicht, so muss man die Modelle von Hand identifizieren. James L. Arbuckle beschreibt das in dem AMOS User Guide Version 3.6 so:

*„...In other applications of common factor analysis, an observed variable can depend on any number of common factors at the same time. In the general case, it can be very difficult to decide whether a common factor analysis model is identified or not. ...If you are unable to tell whether a model is identified, you can try using the model in order to see whether AMOS reports that it is unidentified.“*

## 5 Der konkrete Anwendungsfall

Geprüft werden soll im konkreten Fall die Orthogonalität des Symlograumes. Nach der Reihe werden die, anhand des Adjektiv-Rating-Bogens bewerteten drei Personen untersucht. Die Problematik liegt hier in der Struktur des Raumes und der Anzahl und Struktur der Variablen, welche den Raum darstellen. Der Symlog-Raum soll hier noch einmal kurz grafisch dargestellt werden (Abbildung 7). Wir haben hier drei Dimensionen, d.h. drei Faktoren, die rechtwinklig (d.h. nicht korrelieren) zueinander stehen. Das sind die Faktoren UD, NP und FB. Variablen haben wir insgesamt 26 Stück zur Verfügung. Diese messen entweder für eine Dimension (z.B. U), für zwei Dimensionen (z.B. UF) oder für alle drei Dimensionen (z.B. UFN). Für unser zu prüfendes Modell bedeutet das, dass U auf dem Faktor UD lädt, UF auf den Faktoren UD und FB und UFN auf allen drei Faktoren. Das gilt analog auch für alle anderen Variablen. Die simpelste Lösung, ein 3-Faktorenmodell, ergebe als AMOS-Modell das folgende Bild, welches in Abbildung 3 dargestellt ist.

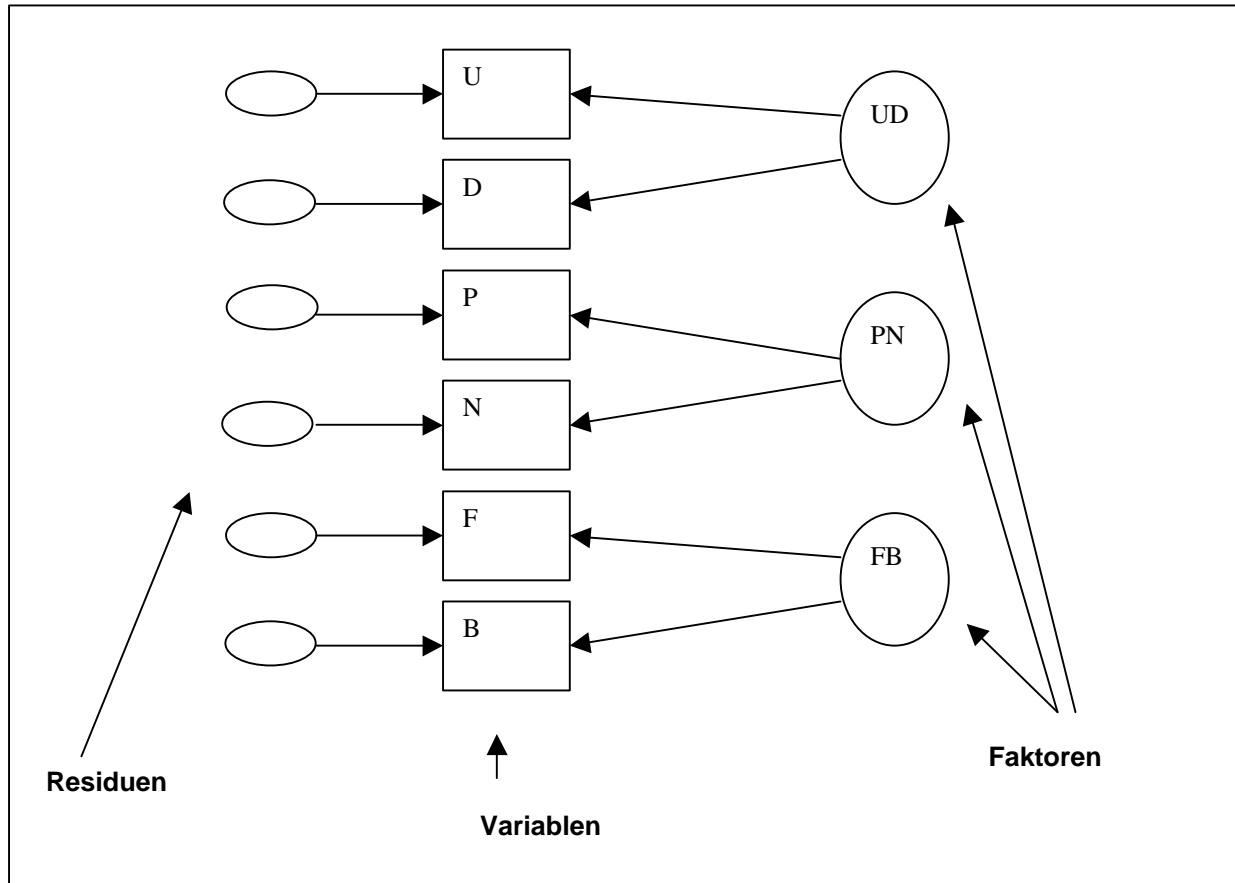


**Abbildung 7:** 3D-Symlog-Raum



Hierzu wurden die Punkte für jede Variable in der U vorkommt zusammengezählt. Also:

$$U + UF + UNF + \dots + UPB = U \text{ etc...}$$



**Abbildung 8:** 3-Faktorenmodell

Ebenso wurden die anderen Variablen berechnet. So ergeben sich die obigen sechs Variablen. Die Variablen U und D laden auf den Faktor UD, denn D ist invers (d.h. liegt im Winkel von 180°) zu U, F ist invers zu B und P invers zu N. Daher die drei Faktoren. Da D das Inverse von U ist, lädt D auf den Faktor UD mit  $-1$ . Entsprechend N und B.

Aber das Problem ist offensichtlich: Auf jeden Faktor laden nur zwei Faktoren. Die Faktoren sind untereinander nicht korreliert. Damit entspricht dieses Modell weder der 3-Indikatorenregel noch der 2-Indikatorenregel. Da die Rechenoperation von der linearen Regression abstammt (Gleichungssysteme), ist eine Identifikation so nicht möglich.

Es gibt nun zwei Möglichkeiten das Problem zu beheben. Die erste Möglichkeit ist die, ein anderes Modell zu suchen, bei dem die Identifikation möglich ist, oder die zweite Möglichkeit ist, weitere Größen zu setzen.

## 5.1 Möglichkeit: Neues Modell erstellen

Beim Symlog-Raum handelt es sich wie gesagt um einen dreidimensionalen Raum, was bedeutet, dass es drei Ebenen gibt, welche unabhängig voneinander sind. Der Adjektiv-Rating-Bogen besteht aus 26 Variablen. Diese sind so aufgeteilt, dass jede Variable neun mal vorkommt, mal in Verbindung mit anderen Variablen mal alleine stehend. Daraus folgt für ein Modell, welches alle Variablen mit einbezieht, dass auf jeden Faktor 18 Variablen laden. Bekannte Parameter sind dann:

- 26 Varianzen
- 325 empirische Kovarianzen  $(25 \cdot (25 + 1)) / 2$ , wobei verschiedene Faktorenkombinationen verwendet werden<sup>7</sup>

Dem gegenüber stehen:

- 54 Faktorladungen der Variablen (18 x 3)
- 26 Varianzen der Meßfehler
- 3 Varianzen der Faktoren

Schon ohne größere Betrachtung dieses Modells wird deutlich, dass es sich hier nicht um ein einfaches Modell handelt, dessen Identifikation Kopfschmerzen bereiten könnte und dies in unserem Fall auch so ist. Trotzdem versuchten wir das Modell berechnen zu lassen, was jedoch bei einem Versuch blieb, den AMOS meldete, das Modell sei nicht identifiziert. Angegeben werden müssten 52 weitere Variablen. (Rechnung 1) Diese Forderung, mit der die Anzahl der Freiheitsgrade (Anzahl der von AMOS zu schätzenden Variablen) erheblich sinken würde, veranlaßte uns dazu auf das einfachere Modell (Abbildung 3) zurückzugreifen und dort die Anzahl der Freiheitsgrade zu senken bzw. zusätzlich zur Identifikation notwendige Modifikationen vorzunehmen.

Auf ein anderes Modell kommen wir später im Rahmen der genaueren Untersuchung zu sprechen.

---

<sup>7</sup> Anhand eines 2-Faktorenmodells sei hier gezeigt, dass nicht alle Kovarianzen zur Identifikation nützlich sind. Gegeben seien 2 voneinander unabhängige Faktoren  $F_1$  und  $F_2$  auf die jeweils 2 Variablen laden. Die Variablen sind von oben durchnummeriert, sodass die erste Variable die 1 trägt. Empirisch sind 6 Kovarianzen vorhanden nämlich KOV(1,2), KOV (1,3), KOV (1,4), KOV (2,3), KOV (2,4) und KOV (3,4). Exemplarisch sei hier eine Gleichung davon herausgegriffen:  $KOV (1,3) = a_1 \times a_3 \times \sqrt{VAR(F_1) \times VAR(F_2)}$ . Durch die Kombination der beiden Faktoren kommt in die Gleichung eine Unbekannte hinzu, sodass diese Gleichung für das Lösen des gesamten Gleichungssystems keine Bedeutung mehr zukommt und deshalb weggelassen werden kann. Also haben wir theoretisch nur noch zwei nützliche Kovarianzen: KOV(1,2) und KOV (3,4). Deshalb ist dieses Modell, wie oben schon erwähnt nicht identifiziert.

## 5.2 Möglichkeit: Zusätzliche Angaben

Wie gesagt kehren wir zum einfachen Modell zurück. Die Gleichungen hierfür lauten:

**Faktor UD:**

$$\text{VAR}(\mathbf{U}) = a_1^2 \times \text{VAR}(f_{UD}) + \text{VAR}(\text{err\_U})$$

$$\text{VAR}(\mathbf{D}) = a_2^2 \times \text{VAR}(f_{UD}) + \text{VAR}(\text{err\_D})$$

$$\text{KOV}(\mathbf{U}, \mathbf{D}) = a_1 \times a_2 \times \text{VAR}(f_{UD})$$

**Faktor PN:**

$$\text{VAR}(\mathbf{P}) = b_1^2 \times \text{VAR}(f_{PN}) + \text{VAR}(\text{err\_P})$$

$$\text{VAR}(\mathbf{N}) = b_2^2 \times \text{VAR}(f_{PN}) + \text{VAR}(\text{err\_N})$$

$$\text{KOV}(\mathbf{P}, \mathbf{N}) = b_1 \times b_2 \times \text{VAR}(f_{PN})$$

**Faktor FB:**

$$\text{VAR}(\mathbf{F}) = c_1^2 \times \text{VAR}(f_{FB}) + \text{VAR}(\text{err\_U})$$

$$\text{VAR}(\mathbf{B}) = c_2^2 \times \text{VAR}(f_{FB}) + \text{VAR}(\text{err\_D})$$

$$\text{KOV}(\mathbf{F}, \mathbf{B}) = c_1 \times c_2 \times \text{VAR}(f_{FB})$$

Zur Identifikation benötigen wir weitere Angaben, welche allerdings begründet werden müssen.

Wir beginnen mit dem Setzen der Faktorvarianzen auf den Wert 1,0, was uns drei zusätzlich bekannte Variablen einbringt. Die Ladungen der Fehler auf die Variablen werden immer auf 1 gesetzt, da man davon ausgeht, dass die gleichen Bedingungen zum Zeitpunkt der Datenerhebung geherrscht haben. Die Fehlereinflüsse sind bei jeder Variablen die gleichen.

Bis hierher haben wir nur routinemäßige Angaben gemacht, die bei jedem Modell gemacht werden. Zur Übersicht halten wir uns die daraus gewonnenen Gleichungen vor Augen:

**Faktor UD**

$$\text{VAR}(\mathbf{U}) = a_1^2 + \text{VAR}(\text{err\_U})$$

$$\text{VAR}(\mathbf{D}) = a_2^2 + \text{VAR}(\text{err\_D})$$

$$\text{KOV}(\mathbf{U}, \mathbf{D}) = a_1 \times a_2$$

**Faktor PN:**

$$\text{VAR(P)} = b_1^2 + \text{VAR(terr\_P)}$$

$$\text{VAR(N)} = b_2^2 + \text{VAR(terr\_N)}$$

$$\text{KOV(P,N)} = b_1 \times b_2$$

**Faktor FB:**

$$\text{VAR(F)} = c_1^2 + \text{VAR(terr\_U)}$$

$$\text{VAR(B)} = c_2^2 + \text{VAR(terr\_D)}$$

$$\text{KOV(F,B)} = c_1 \times c_2$$

Wir haben nun pro Gleichungssystem 3 bekannte und 4 unbekannte Variablen. Hier ist nun der Knackpunkt. Wir brauchen noch eine Angabe pro System, damit das Modell identifiziert ist.

Im folgenden gehen wir davon aus, dass die einzelnen Variablen gleich gut den Sachverhalt messen.<sup>8</sup> Wir setzen also die Koeffizienten  $a_n$ ,  $b_n$  und  $c_n$  der einzelnen Faktoren auf 1 bzw. für Variablen, die theoretisch negativ auf den Faktor laden  $-1$ . Die 1 bedeutet auch in diesem Fall, dass die Variablen perfekt den Sachverhalt widerspiegeln. Zu schätzen bleiben dann noch die Fehlervarianzen. Je größer die Fehlervarianz ist, desto ungenauer ist die Aussage der gesetzten Regressionskoeffizienten.

Betrachten wir den Ausdruck der ersten Berechnung (Rechnung 2).

Wir haben hier den perfekten Fall angenommen, dass die verwendeten Variablen U, D, P, N, F und B alle gleich gut messen und die Faktorenladungen auf 1 bzw.  $-1$  gesetzt. Geschätzt werden können nur noch die Fehlervarianzen. Beim Punkt *Variances* im Ausdruck findet man die Werte der geschätzten Fehlervarianzen. Diese besitzen relativ hohe Werte, was eine unsaubere Erfassung der Daten oder eine praktische Abweichung des Modells von der Theorie bedeutet. Dieses Ergebnis wird durch einen Blick auf verschiedene Masszahlen untermauert. Schon allein das Chi-Quadrat mit einem Wert von 989,950 ist weit davon entfernt nahe bei dem Wert Null zu liegen, was dann bedeuten würde, dass das Modell perfekt den vorhandenen Daten entspricht. Auch die Masszahlen GFI (goodness of fit index; Jöreskog/Sörbom 1986) und AGFI (adjusted goodness of fit index; Jöreskog/Sörbom 1986) zeigen, dass die durch das Modell erklärte Varianz gerade mal bei knapp über 50%, bereinigt um die Freiheitsgrade bei 36,4% liegt. Für diese Masszahlen gilt, dass sie möglichst nahe bei 1 liegen sollen.

<sup>8</sup> Durch die schon oben beschriebene Entstehung (einfache Addition) der Variablen kann man davon ausgehen, dass sogar die Variablen, die auf allen 3 Faktoren laden genauso gut messen, wie eine Variable, die z.B. nur auf den Faktor U lädt. Es ist nicht unsere Arbeit eine vernünftige Kombination der einzelnen Variablen zu finden, sondern zu ergründen, ob eine solche Kombination, wie durch das Symlog-Verfahren gegeben ist, vernünftig ist um die Koordinaten im Symlog-Raum darzustellen.

Noch deutlicher werden diese Aussagen, zieht man DELTA1 (normed fit index; Bentler/Bonett 1980), RHO1 (Bollen 1988) und DELTA2 (modified normed fit index; Bollen 1988) heran. Diese sagen die relative Verbesserung des Modells gegenüber einem Modell, bei dem alle Variablen unabhängig voneinander sind aus. RHO1 versucht DELTA1 um die Freiheitsgrade zu bereinigen. DELTA2 versucht DELTA1 von der Abhängigkeit von der Stichprobengröße und den Freiheitsgraden zu beseitigen. Alle drei Werte sollten nahe bei Eins liegen.<sup>9</sup>

Da diese Variablen den Symlog-Raum nicht darstellen, war von vorne herein zu erwarten, da es sich um ungewichtete und eigentlich völlig willkürlich aufaddierte Variablen handelt. Aus diesem Grund versuchen wir, den Symlog-Raum auf eine andere Art und Weise zu untersuchen.

### 5.3 Ebenenkonstrukt

Der dreidimensionale Raum wird in drei Ebenen aufgeteilt und auf diesen drei Ebenen analysiert. Mit einbezogen werden zur genaueren Untersuchung auch die Variablen, welche auf zwei Faktoren messen (z.B. Variable var\_up). Die Hilfe in dem folgenden Konstrukt liegt in der Schaffung eines neuen Faktors, welcher jeweils im Winkel von 45° zu den Ausgangsfaktoren UD und PN, bzw. UD und FB und FB und PN liegt. Grafisch bedeutet das für beispielsweise die erste Ebene UD und PN folgendes wie in Abbildung 4 dargestellt:

Faktor UD und UP\_ND und Faktor PN und UP\_ND korrelieren jeweils mit 0,5 usw.. Zur vollständigen Beschreibung des Raumes benötigen wir 3 Ebenen welche den Raum rechtwinklig aufspannen. Die Daten zur Berechnung des Modells sind die Originaldaten, d.h. sie sind unverändert, also ohne willkürliche Additionen aufgenommen worden. Als Faktorladungen werden vorgesehen:

Var\_U → Faktor UD

Var\_D → Faktor UD

Var\_P → Faktor PN

Var\_N → Faktor PN

Var\_UP → Faktor PUDN

Var\_UN → Faktor UNPD

Var\_DP → Faktor UNPD

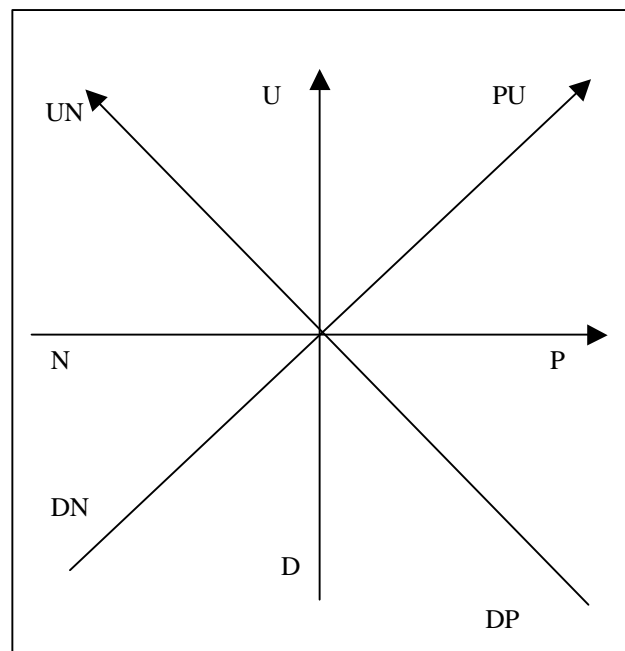


Abbildung 9: Hilfsfaktor PUDN

<sup>9</sup> Auf weitere Ausführungen und Varianten dieses Modells wurde hier verzichtet, da die Ergebnisse für vorgegebene Varianzen der Fehler noch weniger Aussagen machten als das aufgeführte Ergebnis.

Var\_DN → Faktor PUDN

Alle Faktorladungen müssten idealerweise 1.0 bzw –1.0 sein.

Korrelationen zwischen den Faktoren sind in diesem Modell folgende vorhanden:

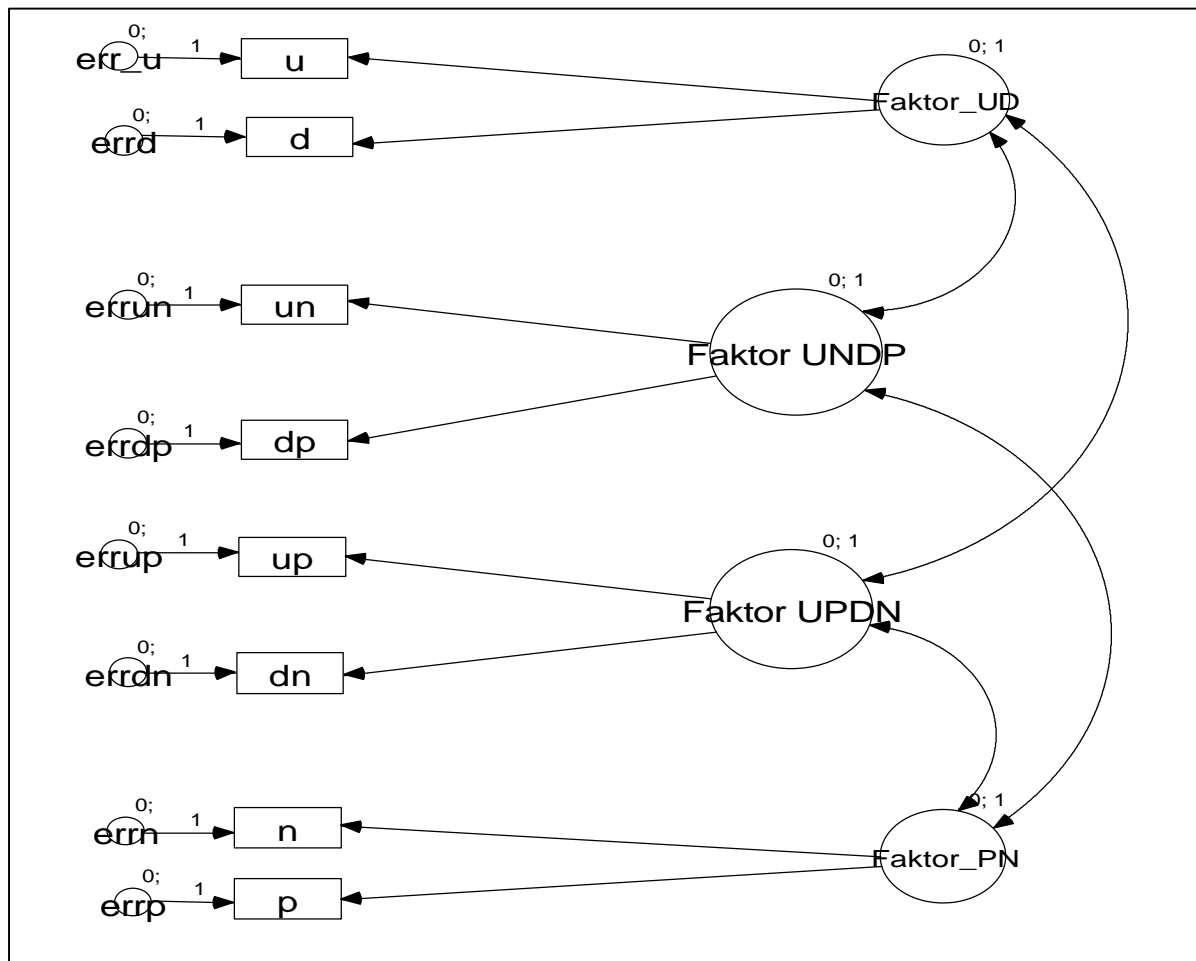
KOV(Faktor UD; Faktor UNPD)

KOV(Faktor PN; Faktor UNPD)

KOV(Faktor UD; Faktor PUDN)

KOV(Faktor PN; Faktor PUDN)

Grafisch dargestellt nimmt das Modell die Formen an, welche in Abbildung 5 dargestellt sind:



**Abbildung 10:** Grafische Darstellung des Modells mit den Faktoren UD und PN

Der grundlegendste Unterschied im Vergleich zu dem Modell in Abschnitt 4 besteht nun darin, dass das Modell schon mit Angabe der Varianzen der Faktoren identifiziert ist. Der Grund dafür liegt im Vorhandensein von Korrelationen zwischen den Faktoren. Dadurch gewinnen wir 4 – 8 Freiheitsgrade hinzu. Zusätzlich befinden sich in diesem Modell auch zweidimensionale Variablen, welche die Berechnungen um einige Potenzen interessanter macht.

Rechnung 3 im Anhang zeigt das Ergebnis<sup>10</sup> der ersten Ebene (UD, PN). Der Chi-Quadrat-Wert ist im Vergleich zum ersten Modell relativ gering. Da dieser aber immer abhängig ist von der Stichprobengröße, kann hier noch keine Aussage gemacht werden. Die wichtigste Bedingung, dass entgegengesetzte Variablen auch tatsächlich entgegengesetzt gemessen werden können. Unter der Teilüberschrift *Regression Weights* sind die unstandardisierten Faktorladungen aufgelistet. U lädt positiv und sein Pendant D negativ. Genauso verhält es sich für die anderen Variablen. Bessere Aussagen erhält man jedoch durch die Betrachtung der *Standardized Regression Weights*. Man erkennt hier, welche Gewichte der Variablen auf den jeweiligen Faktor zukommen. U beispielsweise ist hier mit knapp über der Hälfte seiner Aussagekraft an der Bildung des Faktors UD beteiligt. Die Variable p mit 80% der Aussagekraft an der Bildung von Faktor PN. Im theoretischen Modell sollten die einzelnen Faktorladungen 1.0 bzw. -1.0 sein, was in der Berechnung nur annähernd erfüllt wird. Am schlechtesten lädt die zweidimensionale Variable up, d.h. an dieser Variable müssen bezüglich ihrer Adjektive noch einige Veränderungen vorgenommen werden, dementsprechend auch die Variable dn.

Die Korrelationen zwischen den Faktoren können aufgrund des Zwanges durch die Gleichungssysteme sehr seltsame Werte z.B. über 1 annehmen aber. Im Ergebnis hätten zwei negative Korrelationen stehen müssen – betrachtet man die Theorie. Leider lässt hier das Symlog-Modell oder aber die Art und Weise der Datenerfassung zu wünschen übrig.

Die Maßzahlen hingegen berichten Besseres. DELTA1 und DELTA2 sind beide über 0,95, was soviel bedeutet, dass die Modellanpassung eigentlich gut ist. Auch RHO1 besitzt einen relativ hohen Wert. Betrachtet man allerdings CMIN/DF so ist das Ergebnis nur mittelmäßig. Werte zwischen 2 und 5 repräsentieren eine gute Modellanpassung.<sup>11</sup> Unsere Daten passen also mittelmäßig bis gut auf das oben beschriebene Modell.<sup>12</sup>

Rechnung 4 mit der Ebene, welche durch FB und PN aufgespannt wird, enthält insgesamt schlechtere Ergebnisse wie Ebene UD, PN. Der Chi-Quadrat-Wert ist deutlich höher als bei Rechnung 3 mit der gleichen Anzahl der Fälle in der Stichprobe. Die standardisierten Faktorenladungen für Faktor FB fallen insgesamt, mit Ausnahme von Variable F kleiner aus. Dafür besitzt das Modell die besseren Korrelationen zwischen den Faktoren, die Idealerweise jeweils 0,5 annehmen sollten. DELTA1 und DELTA2 sind wiederum relativ nahe bei 1, ebenso RHO1, jedoch alle niedriger als bei Rechnung 3. Deutlich wird das schlechtere Ergebnis, betrachtet man CMIN/DF. Mit einem Wert von 12,26 entspricht das nicht mehr einer adäquaten Modellanpassung.

<sup>10</sup> Auf das „komplette“ AMOS-Ergebnis wurde hier verzichtet, da z.B. die Titelseite mit dem Namen des Programms und dem Autor keine Rolle spielt.

<sup>11</sup> AMOS Hilfe (Marsh and Hocevar, 1985)

<sup>12</sup> Wir hätten auch ein Modell erstellen können, bei der die Ladungen und Korrelationen vorgegeben sind, hätten dadurch aber eine sehr schlechte Modellanpassung bekommen und hätten dadurch keine Anhaltspunkte um nach Schwächen des Modells zu suchen.

Die oben besprochenen Modelle wurden in erweiterter Form ebenfalls berechnet. Die Faktoren Items u und d müssen theoretisch ebenfalls eine Wirkung auf die Hilfsfaktoren ausüben. Es wurden in einem weitaus komplizierteren Modell deshalb die Hilfsfaktoren mit den vier eindimensionalen Items verbunden. Die berechneten Kennwerte erbrachten wie bei vielen anderen überprüften Modellen keine wesentliche Verbesserung.

Als verbesserungswürdige Variablen seien hier herausgegriffen die Variablen b und nf.

Rechnung 5 mit Ebene FB, PN fehlt, da das Modell in der Form der anderen Modelle nicht identifiziert ist bzw. unseren Rechner bei Iteration Nummer 20342 das gesamte System zum Erliegen gebracht hat.

## 6 Diskussion

Wir haben nun dargelegt, dass sich die Verfahren explorative und konfirmatorische Faktorenanalyse für die Untersuchung des Symlog-Raumes eignen. Gerade bei der explorativen Faktorenanalyse handelt es jedoch um ein sehr datenorientiertes Verfahren. Hier muss deshalb mit erhöhter Vorsicht interpretiert werden. Die Möglichkeit Forschungsartefakte zu erzeugen ist nicht auszuschließen.

„Die erfundene Wirklichkeit“ im Sinne Watzlawicks (1999) stellt eine nicht zu verachtende Kritik an Verfahren wie den hier besprochenen dar. Feyerabend, der sich zugegeben als Wissenschaftsanarchist bezeichnet, stellt in seinem Werk „Wider den Methodenzwang“ komplett die empirische Forschung in Frage. Unter Berücksichtigung dieser Publikationen kann mit Verdrossenheit die Frage nach dem Sinn dieser Untersuchungen gestellt werden.

Weber (1980) versteht jedoch unter Sinn auch eine „durchschnittlich und annähernd in einer gegebenen Masse von Fällen von den Handelnden [...] subjektiv gemeinter Sinn“.

Somit möchten wir die Diskussion so beenden wie Sixtl (1999) sein Werk einleitet: „Wenn die Physiker die Lichtgeschwindigkeit wiederholt messen und das Ergebnis durch folgendes Vertrauensintervall ausdrücken:

$$\Pr(|L - 299.752,5| \leq 0,1) = 0,95$$

dann glauben sie auch daran, daß sie die Lichtgeschwindigkeit „auf hundert Meter genau“ kennen. [...]

Ich will nicht behaupten, daß die Skepsis der Psychologen unbegründet sei. Dagegen behaupte ich, daß sie Ursachen hat, die zum Teil behoben werden können.“



## 7 Fazit

- Verbesserung des Verfahrens mit Hilfe der Faktorenanalyse möglich
- Bestätigung der angenommenen Faktoren
- Weitere Arbeit nötig

Es kann festgehalten werden, dass die Einfachheit der Auswertung mit AMOS, SPSS und Almo in erheblichen Maße abhängig von dem Schwierigkeitsgrad des Modells ist. Die vorhandene Software ermöglicht es einfache Modelle mit einfach, schnell und fehlerfrei zu berechnen. Das eigentliche Problem liegt demnach nicht in der Handhabung von AMOS, SPSS oder Almo, sondern in der Identifikation und der Interpretation der Ergebnisse komplexer Modelle. Die Ergebnisse, die unsere Analysen ergaben, wie das bereits von Krolak-Schwerdt (1994) mit der dreimodalen Faktorenanalyse versucht wurde, „nicht die erwartete Struktur des postulierten dreidimensionalen Symlog Werteraums“. Wir denken aber, dennoch zumindest einige Schwächen des Symlog-Modells aufgedeckt zu haben. Weiterhin möchten wir darauf hinweisen, dass sich die eingesetzten Verfahren zur Analyse dieser Schwächen gut eignen. Betrachtet man gerade die Addition der einzelnen Variablen wie dies im ersten Modell (konfirmatorische Faktorenanalyse) durchgeführt wurde, so konnten wir mit Sicherheit feststellen, dass dieses Vorgehen auf keinen Fall der empirischen Realität des Datensatzes entspricht.

- Die Variablen dn, up, nf und d sollten aufgrund unseres 3-Ebenen-Modellen erneut in ihrer Operationalisierung überdacht werden.

Dennoch muss man mit Respekt anerkennen, dass das Symlog-Verfahren, tatsächlich einen Raum vektorisiert, welcher auf einer Reihe von Adjektiven basiert. Das allein ist schon sehr erstaunlich, wenn man bedenkt wie viele unterschiedliche Interpretationen es hinsichtlich der verwendeten Adjektive gibt, welche Fehler durch die Stichprobe geschulter und ungeschulter Rater und welche Verzerrungen durch die geringe Anzahl der beobachteten Personen auftreten. Außerdem ist es mit diesem Verfahren trotz der Unschärfe der Variablen möglich beobachtete Personen zumindest grob zu charakterisieren.

Wir haben somit das gesetzte Ziel, die theoretische Auseinandersetzung mit den besprochenen Verfahren und die Anwendung an einem konkreten Datensatz, vollends erreicht. Darüber hinaus können wir wahrscheinlich aufgrund unserer Analysen den Erkenntnisgewinn in der Kleingruppenforschung vorantreiben.

## 8 Literatur

**AMOS User Help (1994-1997)**, Version 3.6, SmallWaters Cooperation,

**AMOS (1998)**: User Help, Version 4.01, SmallWaters Cooperation,

**Arbuckle, James L. (1997)**: „AMOS Users' Guide Version 3.6“, 1997, SPSS Inc.

**Bacher, Johann (1996)**: „Clusteranalyse: anwendungsorientierte Einführung“. – 2., ergänzte Ausgabe – München, Wien: Oldenburg Verlag.

**Bacher, Johann (WS 99/00)**: „Skalierung und Clusteranalyse - Vorlesungsskript, WS 99/00.“

**Bales, F., Robert und Cohen, P., Stephen (1982)**: „SYMLOG: e. System für d. mehrstufige Beobachtung von Gruppen“ – Aus dem Amerikanischen übers. Von Johann Schneider und Peter Orlik – Stuttgart: Klett-Cotta.

**Borg, Ingwer und Staufenbiel, Thomas (1997)**: „Theorien und Methoden der Skalierung: eine Einführung“ – 3., überarb. Aufl. – Bern; Göttingen; Totonto; Seattle: Hans Huber Verlag. (Methoden der Psychologie; Bd. 11)

**Bortz, Jürgen (1999)**: „Statistik für Sozialwissenschaftler“. – 5., vollst. überarb. Aufl. – Berlin, Heidelberg, u.a.: Springer Verlag.

**Feyerabend, Paul K. (1991)**: „Wider den Methodenzwang“. - 3. Aufl. Frankfurt am Main: Suhrkamp

**Forum für Strukturgleichungsmodelle**

<http://www.gsu.edu/~mkteer/semnet.html> Zugriffen am 26.04.00

**Geider, Franz J.; Rogge, Klaus-Eckart; Schaaf, Harald P. (1982)**: „Einstieg in die Faktorenanalyse“. Heidelberg: Quelle & Meyer

**Holm, Kurt (1998)**: „ALMO Statistik-System Handbuch – P30. Faktorenanalyse“. Linz.

**Krolak-Schwerdt, Sabine (1994)**: „Ein Methodenvorschlag für interkulturelle Vergleiche“. In: Gruppendynamik Heft 2, Juni 1994.

**Leibniz Rechenzentrum München**

<http://www.lrz-muenchen.de/services/software/statistik/amos> Zugriffen am 26.04.00

**Pawlik, Kurt (1976)**: „Dimensionen des Veraltens – Eine Einführung in Methodik und Ergebnisse faktorenanalytischer psychologischer Forschung“. – 3., unveränderte Auflage – Bern; Stuttgart; Wien: Hans Huber Verlag.

**Roth, Erwin (1984)**: „Sozialwissenschaftliche Methoden“, 1984, Oldenbourg

**Seber, G. A. F. (1938)**: „Multivariate observations“ : John Wiley & Sons, Inc.

**Sixtl, Friedrich (1996)**: „Einführung in die Exakte Psychologie“. – München, Wien: Oldenburg

**Tabachnik, Barbara und Fidell, S., Linda (1996)**: „Using Multivariate Statistics“. - Third Edition – New York: HarperCollins College Publishers.

**Überla, K (1977)**: Faktorenanalyse – Eine systematische Einführung für Psychologen, Mediziner, Wirtschafts- und Sozialwissenschaftler“. – Nachdruck der zweiten Auflage – Berlin; Heidelberg; New York: Springer Verlag

**Watzlavik, Paul (1999)**: „Die erfundene Wirklichkeit“. 11.. Auflage – München: Piper.

**Weber, Max (1980)**: „Wirtschaft und Gesellschaft: Grundriß d. verstehenden Soziologie“. – 5., rev. Aufl., Studienausg. – Tübingen: Mohr.

## 9. Abbildungsverzeichnis

- Abbildung 1:** Faktorenmodell (Überla)  
**Abbildung 2:** Veranschaulichung verschiedener Korrelationen (Borg)  
**Abbildung 3:** Stricknadelmodell (Borg)  
**Abbildung 4:** Vektorenkonfiguration (Borg)  
**Abbildung 5:** Der Symlograum (aus Bales)  
**Abbildung 6:** Einfaches Beispiel  
**Abbildung 7:** 3D-Symlog-Raum  
**Abbildung 8:** 3-Faktorenmodell  
**Abbildung 9:** Hilfsfaktor PUDN  
**Abbildung 10:** Grafische Darstellung des Modells mit den Faktoren UD und PN

## 10. Tabellenverzeichnis

- Tabelle 1:** Korrelationsmatrix (Borg)  
**Tabelle 2:** Hauptkomponentenanalyse  
**Tabelle 3:** Image-Faktorisierung  
**Tabelle 4:** Oblimin-Lösungen  
**Tabelle 5:** Komponentenkorrelationsmatrix  
**Tabelle 6:** Rotierte Komponentenmatrix  
**Tabelle 7:** Rotierte Komponentenmatrix  
**Tabelle 8:** Fünffaktorenlösung  
**Tabelle 9:** Identifikationsregeln

**Der hier verwendete Datensatz steht bei Christine Marx Lehrstuhl Kreutz Uni Erlangen-Nürnberg jederzeit zur Verfügung**

# Anhang

**Rechnung 1**

**Rechnung 2**

**Rechnung 3**

**Rechnung 4**