

# Faktorenanalyse

Auf den nächsten Seiten soll das Verfahren der Faktorenanalyse an einigen Beispielen dargestellt werden. Ziel ist es, die Durchführung von explorativen Faktorenanalysen mit SPSS für Windows zu erläutern und insbesondere die inhaltliche Interpretation von Ergebnissen der Faktorenanalyse zu erleichtern. Die Ausführungen zum theoretischen Hintergrund der verschiedenen Verfahren und zu den statistischen Maßzahlen fallen dementsprechend knapp aus - der Rückgriff auf einschlägige Fachliteratur soll hier auf keinen Fall ersetzt, sondern angeregt werden.

## 1. Grundsätzliche Überlegungen<sup>1</sup>

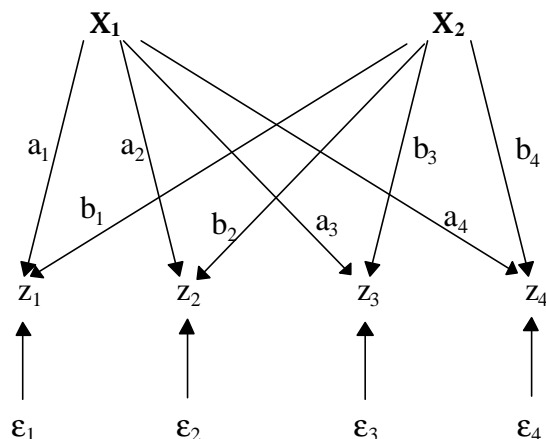
Allgemein ausgedrückt ist es das Ziel einer Faktorenanalyse, die Struktur einer Korrelationsmatrix (oder auch einer Kovarianzmatrix) zu "erklären" - d.h. latente Variablen (Faktoren) zu generieren, deren Zusammenspiel ein möglichst genaues Abbild der Ausgangsmatrix erzeugt. Dabei können grundsätzlich zwei Herangehensweisen unterschieden werden:

- Die explorative Faktorenanalyse. Hier wird es (mehr oder weniger) dem mathematischen Algorithmus überlassen, wieviele interpretierbare Faktoren sich aus einer Korrelationsmatrix generieren lassen (Faktorenanalyse mit SPSS). Annahmen über die Struktur der Faktorladungen werden nicht gemacht. Im folgenden besprechen wir ausschließlich die Hauptkomponentenmethode.
- Die konfirmatorische Faktorenanalyse. Hier werden vor Anwendung des Verfahrens Hypothesen bezüglich der zugrundeliegenden Faktoren formuliert. Es wird dann im Rahmen der Datenanalyse untersucht, wie gut sich die gegebene Korrelationsmatrix mit Hilfe des vorgegebenen Faktorstrukturmodells reproduzieren läßt (LISREL). Eine LISREL-Analyse ist ebenfalls mit SPSS für Windows möglich, wird aber im folgenden nicht näher erläutert.

Um die theoretischen Grundgedanken der Faktorenanalyse zu verstehen, ist der kausalanalytische Ansatz sehr hilfreich. Zur Illustration verwenden wir ein Modell mit zwei Faktoren und vier Indikatoren; die abgeleiteten Formeln können dabei jedoch ohne Schwierigkeiten auf den allgemeinen Fall erweitert werden ( $m$  Faktoren,  $n$  Variablen,  $m < n$ ).

---

<sup>1</sup> Wichtige Erläuterungen zur Faktorenanalyse befinden sich auch im Skript zur Vorlesung „Methoden der empirischen Wirtschafts- und Sozialforschung“, Abschnitt 5.4.2.



Faktoren sind theoretische Konstrukte ( $\xi_1$  und  $\xi_2$ ), denen *Wirkungsindikatoren* ( $z_1, z_2, z_3, z_4$ ) zugeordnet werden. Die *Faktorladungen* ( $a_1 \dots a_4, b_1 \dots b_4$ ) entsprechen den Korrelationen der latenten Variablen (Faktoren) mit den Indikatoren. Dabei wird vorausgesetzt, daß die beiden Faktoren voneinander unabhängig (*orthogonal*) sind, also keine Korrelation miteinander aufweisen. (Diese Annahme kann jedoch bei bestimmten Verfahren aufgehoben werden, z.B. im Falle einer schiefwinkligen Rotation). Die  $\epsilon_i$  fassen spezifische Einflüsse und Zufallsfehler zusammen, die jeweils auf die Variable  $z_i$  wirken und die Einzelrestvarianz bestimmen. Die  $\epsilon_i$  sind untereinander und mit den latenten Faktorvariablen unkorreliert.

Formal wird dieses Modell durch folgende *Strukturgleichungen* ausgedrückt:

$$z_1 = a_1 \xi_1 + b_1 \xi_2 + \epsilon_1$$

$$z_2 = a_2 \xi_1 + b_2 \xi_2 + \epsilon_2$$

$$z_3 = a_3 \xi_1 + b_3 \xi_2 + \epsilon_3$$

$$z_4 = a_4 \xi_1 + b_4 \xi_2 + \epsilon_4$$

Die *Korrelation* zwischen zwei Indikatoren läßt sich für dieses Modell nach folgender Formel berechnen:<sup>2</sup>

$$r_{z_i z_j} = a_i a_j + b_i b_j$$

Da die Faktoren voneinander unabhängig sind, ergibt sich die durch die Faktoren *erklärte Varianz*  $h_i^2$  (auch gemeinsame Varianz genannt) für jeden Indikator  $i$  als Summe der quadrierten Faktorladungen:

$$h_i^2 = a_i^2 + b_i^2$$

---

<sup>2</sup> Diese Formel läßt sich relativ einfach herleiten. Da nach Voraussetzung die Variablen standardisiert und die Fehlerterme unkorreliert sind, ergibt sich die Formel für die Korrelation zwischen  $z_i$  und  $z_j$  nach den Regeln der Pfadanalyse als Summe zweier scheinkorrelativer Effekte (vgl. Skript 3.1.1.3, b) Typologie kausaler Effekte)

Die gemeinsame Varianz  $h^2$  eines Indikators  $i$  wird auch als *Kommunalität* bezeichnet. Da von standardisierten Variablen ausgegangen wird, ist die Gesamtvarianz gleich eins; sie kann wie folgt in die gemeinsame Varianz (Kommunalität) und die Einzelrestvarianz zerlegt werden:

$$\text{Var}(z_i) = 1 = h_i^2 + u_i^2 \quad [u_i^2 \text{ ist die Einzelrestvarianz (durch gemeinsame Faktoren nicht erklärte Varianz), die auf } \varepsilon_i \text{ zurückzuführen ist}]$$

Sumiert man die quadrierten Faktorladungen eines Faktors über alle Indikatoren auf, so erhält man den Teil der Gesamtvarianz aller Indikatoren, der durch diesen Faktor erklärt wird. In der unrotierten Lösung einer Hauptkomponentenanalyse entspricht dieser Anteil dem Eigenwert des Faktors. Zur Interpretation dieses Wertes muß man sich verdeutlichen, daß von standardisierten Indikatoren ausgegangen wird, die jeweils eine Varianz von 1 aufweisen. Die Gesamtvarianz entspricht also der Anzahl der Indikatoren  $N$ . Um den prozentualen Anteil an erklärter Gesamtvarianz zu erhalten, muß deshalb die Summe der quadrierten Faktorladungen (bzw. der zugehörige Eigenwert) eines Faktors durch  $N$  geteilt werden.

## 2. (hypothetisches) Beispiel: Schulnoten

Wir untersuchen die folgende (hypothetische) Korrelationsmatrix, welche die Korrelationen zwischen den Schulnoten von Schülern in den Fächern Deutsch (D), Englisch (E), Physik (P) und Mathematik (M) enthält (vgl. Kapitel 5.4.2 im Skript):

Correlation Matrix:

	D	E	P	M
D	1,00000			
E	,58000	1,00000		
P	,25000	,30000	1,00000	
M	,17000	,25000	,74000	1,00000

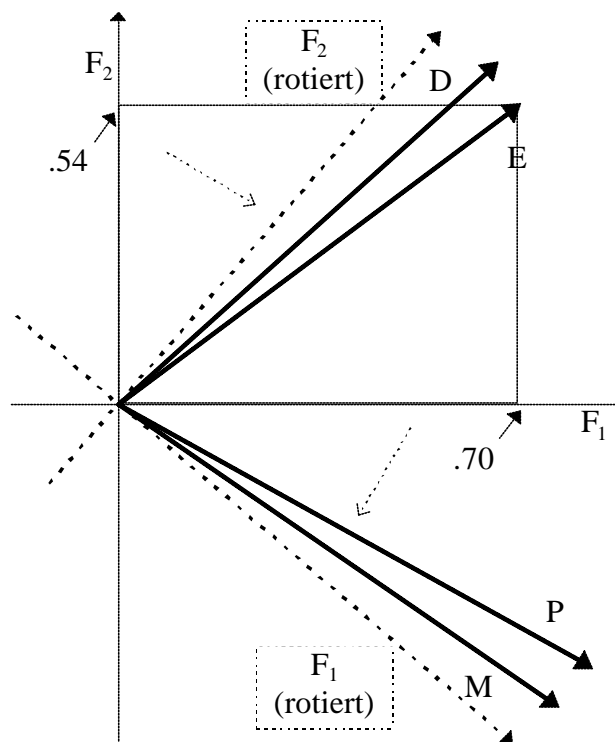
Alle Fachnoten korrelieren positiv untereinander, es fällt jedoch auf, daß die beiden Fachnoten in Deutsch und Englisch, ebenso wie die beiden Noten in Physik und Mathematik, in einem engeren Zusammenhang zueinander stehen als zu den übrigen Noten. Geht man davon aus, daß zwei Faktoren - mathematische und sprachliche Bgabung - für die Schulnoten in diesen vier Fächern verantwortlich sind, so ist diese Korrelationsstruktur bereits ein Hinweis auf die Konvergenz der Indikatoren zum gleichen theoretischen Konstrukt [Faktor] und die Diskriminanz der Indikatoren von verschiedenen theoretischen Konstrukten [Faktoren] (vgl. Schnell/Hill/Esser, Kap. 4: Konstruktvalidität). Bei einer Faktorenanalyse nach der Hauptkomponentenmethode werden tatsächlich zwei Faktoren extrahiert:

	Factor 1	Factor 2
D	,64148	,62593
E	,70038	,53907
P	,81362	-,45162
M	,76804	-,53594

Bildet man die Summe der quadrierten Faktorladungen für jeden Faktor, so erhält einen Betrag von 2,154 für den ersten Faktor und einen Betrag von 1,174 für den zweiten Faktor. Die Gesamtvarianz aller Indikatoren beträgt in diesem Beispiel 4. Somit erklärt der erste Faktor  $2,154/4 = 53,8\%$  an Varianz, der zweite Faktor erklärt  $1,174/4 = 29,3\%$  der Varianz der Indikatoren. Insgesamt werden also  $83,2\%$  der Varianz der Indikatoren durch die beiden Faktoren erklärt. Die erklärte Varianz der einzelnen Indikatoren durch beide Faktoren (Kommunalität) ist in der folgenden Tabelle abzulesen:

Variable	Communality
D	,80328
E	,78112
P	,86595
M	,87711

Die Ergebnisse sind recht gut, allerdings ist die Lösung der Faktorenanalyse in der obigen Form noch nicht sehr aussagekräftig. Die folgende Grafik verdeutlicht das Ergebnis der Faktorenanalyse.



Jeder Indikator wird hier durch einen Pfeil im Koordinatensystem dargestellt. Die Faktorladungen lassen sich für jeden Indikator am Koordinatensystem (bestehend aus Faktor 1 und 2) ablesen. Die Länge der Pfeile entspricht der Kommunalität des Indikators.<sup>3</sup> In der ursprünglichen Lösung (Koordinatensystem mit durchgezogenen Linien) laden alle Indi-

<sup>3</sup> Aus technischen Gründen stimmen die Werte in der Grafik nur näherungsweise.

katoren positiv auf Faktor 1. Die Indikatoren P und M laden negativ auf Faktor 2, die Ladungen für D und E sind hier positiv.

Die Ausgangsbedingung für die Faktorenanalyse schreibt jedoch (abgesehen von der Orthogonalität der Faktoren) die Lage des Koordinatensystems nicht vor. Es ist also möglich, die Koordinaten (= Faktoren) so zu rotieren, daß die Indikatoren jeweils auf einem Faktor eine sehr hohe Ladung, auf dem anderen Faktor dagegen eine möglichst geringe Ladung aufweisen (sog. VARIMAX-Rotation). Diese Rotation ist durch die gestrichelten Linien angedeutet (beide Faktoren wurden um ca. 45 Grad im Uhrzeigersinn gedreht). Die Faktorladungen bezüglich dieses rotierten Koordinatensystems stellen die rotierte Faktorladungsmatrix dar:

Rotated Factor Matrix:

	Factor 1	Factor 2
D	,08138	,89256
E	,18258	,86475
P	,91201	,18488
M	,93209	,09116

Zunächst ist festzuhalten, daß die Kommunalitäten der Indikatoren sich nach der Rotation nicht geändert hat - die Summe der quadrierten Faktorladungen ist für jeden Indikator nach der Rotation genauso groß wie vor der Rotation. Auch die durch beide Faktoren erklärte Varianz ist mit 83,2% gleichgeblieben. Geändert hat sich allerdings der Anteil an der Gesamtvarianz, der von den einzelnen Faktoren erklärt wird. Die Summe der quadrierten Faktorladungen für Faktor 1 ergibt nun 1,74. Dies entspricht 43,5% der Gesamtvarianz. Die Summe der quadrierten Faktorladungen des zweiten Faktors beträgt 1,59. Faktor 2 erklärt 39,7% der Gesamtvarianz.

Die rotierte Lösung läßt sich in diesem Fall sehr gut inhaltlich interpretieren: Die Schulnoten der Fächer Physik und Mathematik sind Indikatoren für die *mathematische Begabung* (Faktor 1) der Schüler, die Noten in Deutsch und Englisch können als Indikatoren der *sprachlichen Begabung* (Faktor 2) der Schüler interpretiert werden.

Ein Vergleich mit den wahren Faktorwerten, die zur Konstruktion der Korrelationsmatrix zwischen den Schulnoten verwendet wurden (vgl. Kap. 5.4.2 im Skript) zeigt zwar eine grobe Übereinstimmung aber auch einige Abweichungen. Generell gilt, daß bei einer Hauptkomponentenanalyse die Kommunalitäten überschätzt werden, wie dies auch der Vergleich mit den wahren Kommunalitäten in unserem Beispiel bestätigt (vgl. Kap. 5.4.2 im Skript).<sup>4</sup>

---

<sup>4</sup> Als Alternative zur Hauptkomponentenanalyse wäre eine Hauptachsenanalyse in Erwägung zu ziehen, die die Kommunalitäten iterativ schätzt. Allerdings ist dieser „Vorteil“ der Hauptachsenanalyse auch zugleich das Problem, da eine eindeutige Begründung für die Vorgehensweise bei der Kommunalitätenschätzung fehlt. Für einen „erfahrenen Analytiker“ stellt aber der Vergleich einer Hauptkomponenten- und einer Hauptachsenanalyse eine wertvolle Entscheidungshilfe für eine abschließende Interpretation der Faktorstruktur dar.

### 3. Beispiel: Messung des allgemeinen Selbstwertes

Als Beispiel aus einer empirischen Untersuchung erläutern wir die Vorgehensweise bei einer explorativen Faktorenanalyse zunächst an einer Likertskala<sup>5</sup> zur Messung des *allgemeinen Selbstwertes*.<sup>6</sup> Diese Skala entstammt der Konsumsuchtstudie von Gerhard Scherhorn. Der Selbstwert umfaßt Einstellungen (Attitüden) eines Individuums gegenüber sich selbst - sehr allgemein ausgedrückt: das „Selbstbewußtsein“ einer Person (vgl. Scherhorn et al 1993, S. 28ff)<sup>7</sup>. Den Befragten wurden 10 Items vorgelegt, zu denen sie auf einer 6-stufigen Ratingskala Zustimmung oder Ablehnung äußern konnten:

Item	Wortlaut	Nr. im Datensatz
Item 1	Manchmal glaube ich, daß ich zu nichts gut bin	V226n
Item 2	Ich finde mich ganz in Ordnung	V227
Item 3	Ich bin ein Niemand	V228n
Item 4	Eigentlich bin ich mit mir ganz zufrieden	V229
Item 5	Manchmal wünschte ich, ich wäre nicht geboren	V230n
Item 6	Ich wollte, ich könnte mehr Achtung vor mir haben	V231n
Item 7	Manchmal fühle ich mich zu nichts nütze	V232n
Item 8	Wenn ich mich mit anderen Menschen meines Alters vergleiche, schneide ich eigentlich ganz gut ab	V233
Item 9	Ich verachte mich	V234n
Item 10	Ich bin zufrieden mit mir	V235

Ein hohes Rating der Items signalisiert Zustimmung. Die Items 1,3,5,6,7 und 9 sind also im Sinne der Messung von Selbstwert negativ gepolt. Ein hoher Wert (Zustimmung!) zeigt bei diesen Items einen geringen Selbstwert des/der Befragten an. Für die weitere Analyse wurden diese Items umgepolt (z.B. V226n = 7-V226).

---

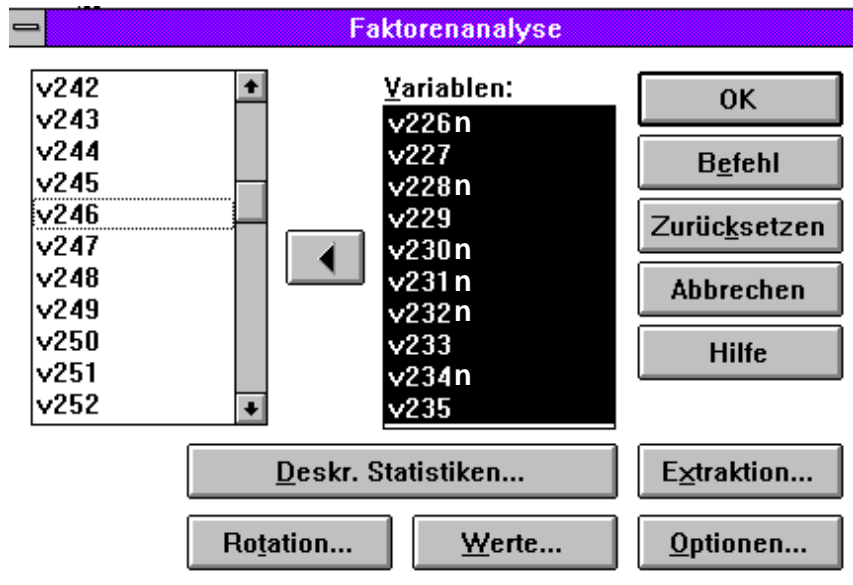
<sup>5</sup> vgl. Kapitel 5.3.1 im Skript zur Vorlesung „Methoden der empirischen Wirtschafts- und Sozialforschung“.

<sup>6</sup> Zu der folgenden Faktorenanalyse vgl. auch Kap. 5.4.2 im Skript.

<sup>7</sup> Scherhorn, Gerhard; Ariane Gläser; Michael Neuner; Gerhard Raab; Lucia Reisch (1993): *Indikatoren der Bevölkerungsumfrage*, Lehrstuhl für Konsumtheorie und Verbraucherpolitik Universität Hohenheim, Arbeitspapier 62, Stuttgart.

### 3.1 Die explorative Faktorenanalyse mit SPSS für Windows

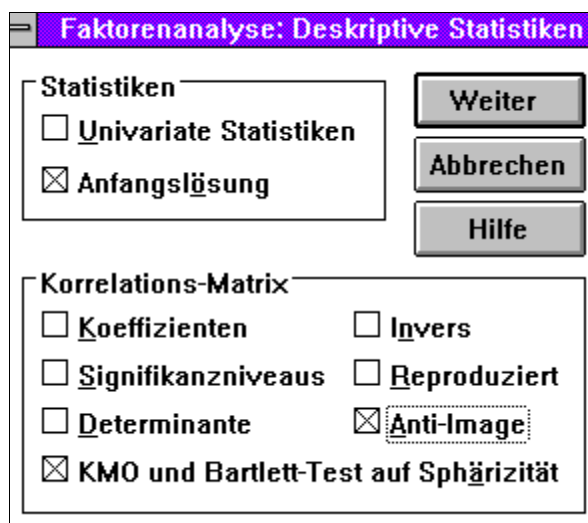
Nachdem der Datensatz der Scherhornstudie in SPSS geladen wurde, wählen wir unter dem Auswahlm Menü **Statistiken** das Submenü **Datenreduktion** und hier die **Faktorenanalyse**. Es erscheint das folgende Dialogfenster:



Wir übertragen die entsprechenden Variablen in das Variablenfenster.

Zusätzlich können wir die Prozedur über die weiteren Dialogfelder anpassen ...

Das Dialogfeld **Deskriptive Statistiken** enthält folgende Auswahlmöglichkeiten:



Auf jeden Fall sollte hier die **Anfangslösung** angefordert werden.

Daneben fordern wir in unserem Beispiel noch den **KMO und Bartlett-Test** sowie die **Anti-Image** Korrelationsmatrix an (zu diesen Statistiken später mehr)

Hinter dem Dialogfeld **Extraktion** verbirgt sich folgendes Dialogfenster:

**Faktorenanalyse: Extraktion**

Methode: **Hauptkomponenten**

**Extrahieren**

**Eigenwerte über:**

**Anzahl Faktoren:**

**Anzeigen**

**Unrotierte Faktorenlösung**

**Screplot**

**Maximale Iterationen für Konvergenz:**

Weiter  
Abbrechen  
Hilfe

Unter **Methode** stehen verschiedene *Schätzverfahren* zur Verfügung. Die gängigsten Verfahren sind die *Hauptkomponentenanalyse* und die *Hauptachsenanalyse*. Liegt keine spezifische Hypothese über die Größenordnung der Kommunalitäten vor, so ist eine Hauptkomponentenanalyse vorzuziehen, da sie eine mathematisch exakte und methodisch nachvollziehbare Zerlegung der gesamten Varianz aller Indikatoren über eine Eigenwert-Analyse der Korrelationsmatrix darstellt. Dabei entspricht die Summe aller N Eigenwerte genau der Gesamtvarianz aller Indikatoren, also N (siehe unten).

Unter dem Stichwort **Extrahieren** kann ausgewählt werden, wieviele Faktoren bei der Analyse extrahiert werden sollen. In der Regel wird hier die Option **Eigenwerte über 1** gewählt. Dies ist das sog. *Kaiser-Kriterium*, welches besagt, daß ein Faktor zumindest mehr als die Varianz eines einzigen Items erklären soll. Alternativ kann jedoch auch die **Anzahl** der zu extrahierenden Faktoren vorgegeben werden. Desweiteren kann hier neben der **unrotierten Faktorenlösung** auch ein sog. **Screplot** angefordert werden. Im Screplot werden die Eigenwerte der Faktoren grafisch dargestellt, er dient als Entscheidungshilfe bei der Frage, wieviele Faktoren für eine optimale Lösung extrahiert werden sollten.

Um die rotierte Lösung (bei mehrfaktoriellen Ergebnissen) anzufordern, muß das Dialogfeld **Rotation** gewählt werden:

**Faktorenanalyse: Rotation**

**Methode**

**Keine**  **Quartimax**

**Varimax**  **Direktes Oblimin**

**Equamax** **Delta:**

**Anzeigen**

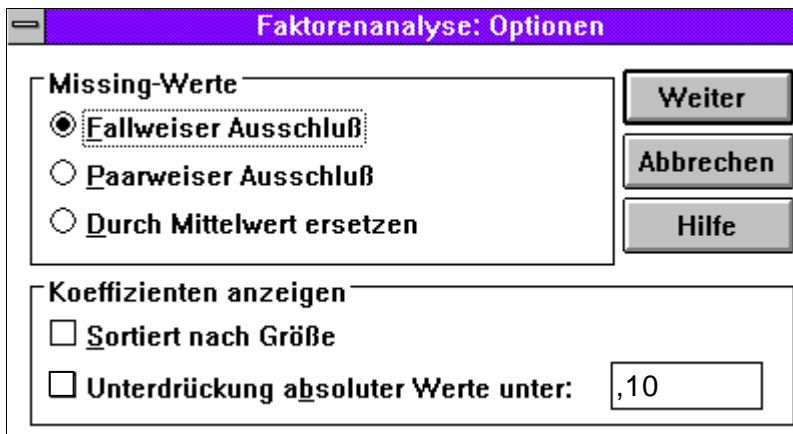
**Rotierte Lösung**  **Ladungsplot(s)**

**Maximale Iterationen für Konvergenz:**

Das in den meisten Fällen sinnvollste Verfahren ist die **Varimax**-Rotation. Sofern die Annahme unkorrelierter Faktoren aufgegeben wird, kann hier auch eine schiefwinklige (oblique) Rotation durch Auswahl von **Direktes Oblimin** gewählt werden. Auf jeden Fall sollte das Kästchen **Rotierte Lösung** angekreuzt werden.

Desweiteren können folgende **Optionen** gewählt werden:





Der **Fallweise Ausschluß** der **Missing-Werte** ist bei der Faktorenanalyse meistens zu empfehlen, da nur so sichergestellt werden kann, daß die zugrundeliegende Korrelationsmatrix tatsächlich auch faktoranalytisch untersucht werden kann.<sup>8</sup>

Die **Unterdrückung absoluter Werte unter** einem frei wählbaren Betrag macht den Output der Faktorenanalyse wesentlich übersichtlicher. Außerdem können die Ergebnisse auch **Sortiert nach Größe** ausgegeben werden. In diesem Fall werden die zu einem Faktor gehörenden Items von den übrigen Items abgesetzt.

Das hier nicht näher dargestellte Dialogfeld **Werte** ermöglicht die Erzeugung neuer Variablen, welche die Faktorscores der Faktoren enthalten. In vielen Fällen ist es jedoch sinnvoller, die Items jedes Faktors „per Hand“ zu Subskalen aufzusumieren.

### 3.2 Faktorenanalyse der Skala „allgemeiner Selbstwert“

Die unter 3.1 getroffenen Einstellungen erzeugen den folgenden SPSS-Befehl:<sup>9</sup>

```
FACTOR
/VARIABLES v226n v227 v228n v229 v230n v231n v232n v233 v234n v235 /MISSING
LISTWISE /ANALYSIS v226n v227 v228n v229 v230n v231n v232n v233 v234n v235
/PRINT INITIAL KMO AIC EXTRACTION ROTATION
/CRITERIA MINEIGEN(1) ITERATE(25)
/EXTRACTION PC
/CRITERIA ITERATE(25)
/ROTATION VARIMAX .
```

Dieser Befehl liefert folgenden Output:

```
----- FACTOR ANALYSIS -----
Analysis number 1 Listwise deletion of cases with missing values
Kaiser-Meyer-Olkin Measure of Sampling Adequacy = ,89455
Bartlett Test of Sphericity = 5404,0713, Significance = ,00000
```

Zunächst werden das KMO-Maß sowie der Bartlett Test of Sphericity ausgegeben. Das KMO-Maß (Kaiser-Mayer-Olkin Measure) gibt Auskunft über die Güte der Fakto-

<sup>8</sup> Die Matrix muß positiv definit sein.

<sup>9</sup> Die Befehle können durch die Auswahl des Dialogfeldes **Befehl** anstelle von **OK** in ein Syntaxfenster übertragen werden.

renanalyse. Ein KMO-Wert über .89 deutet auf eine besonders gute Eignung der Items für eine Faktorenanalyse hin. Der Bartlett-Test ist signifikant. Das bedeutet, daß die *Nullhypothese* dieses Tests, die Korrelationsmatrix sei nur zufällig von der Einheitsmatrix verschieden, mit einer Irrtumswahrscheinlichkeit von nahezu 0,000% abgelehnt werden kann. (Zur Interpretation beider Maße vgl. z.B. Brosius 1995, S. 821ff).

Es folgt die Anti-image Covariance Matrix, die hier nicht näher erläutert werden soll.

Anti-image Covariance Matrix:  
[ ... gelöscht ... ]

Anschließend wird die Anti-image Correlation Matrix ausgegeben:

Anti-image Correlation Matrix:

	V226n	V227	V228n	V229	V230n	V231n	V232n
V226n	,90740						
V227	-,14045	,88493					
V228n	-,21949	-,05965	,90918				
V229	,02047	-,27106	,01384	,85597			
V230n	-,08997	-,00332	-,12912	-,08360	,90976		
V231n	-,12399	-,05151	-,08991	,01798	-,12214	,92671	
V232n	-,26022	,02480	-,05536	-,15511	-,20851	-,23884	,89834
V233	-,01900	-,19946	-,04379	-,13361	-,06395	-,05662	-,01114
V234n	-,02525	,02353	-,27093	,00196	-,28194	-,12428	-,11712
V235	-,00411	-,12925	-,02051	-,33457	,01534	-,00430	-,01877

	V233	V234n	V235
V233	,89300		
V234n	,05922	,88667	
V235	-,23050	-,09807	,86306

Measures of Sampling Adequacy (MSA) are printed on the diagonal.

Diese Matrix enthält (unterhalb der Diagonalen) die negativen Werte der partiellen Korrelationen zwischen den Items, die sich ergeben, wenn jeweils alle anderen Items kontrolliert werden. Im Idealfall würden hier nur noch Korrelationen mit einem Betrag von Null übrigbleiben (was in empirischen Untersuchungen naturgemäß nie exakt eintreffen wird). Aber auch in diesem Beispiel gibt es nur wenige wirklich erwähnenswert hohe Restkorrelationen. Die partielle Korrelation zwischen V229 und V235 ist hier mit .33457 am höchsten. Offenbar gibt es unabhängig vom allgemeinen Selbstwert noch einen positiven Zusammenhang zwischen der Beantwortung der Frage, ob ich „mit mir zufrieden“ bin (V229) oder „im Vergleich mit anderen“ gut abschneide (V233).

Die Diagonale dieser Matrix enthält die sog. MSA-Werte (measure of sampling adequacy), die sich inhaltlich ähnlich interpretieren lassen, wie das KMO-Maß (vgl. Brosius 1995, S. 823), allerdings auf der Ebene der einzelnen Items. Alle MSA-Werte liegen über .85, teilweise sogar über .90, es liegt hier also kein Hinweis vor, ein oder mehrere Items von der Faktorenanalyse auszuschließen.

Es folgt die Ausgabe der Initial Statistics:<sup>10</sup>

Extraction 1 for analysis 1, Principal Components Analysis (PC)

Initial Statistics:

Variable	Communality	*	Factor	Eigenvalue	Pct of Var	Cum Pct
V226n	1,00000	*	1	4,52500	45,2	45,2
V227	1,00000	*	2	1,41098	14,1	59,4
V228n	1,00000	*	3	,66128	6,6	66,0
V229	1,00000	*	4	,60190	6,0	72,0
V230n	1,00000	*	5	,59331	5,9	77,9
V231n	1,00000	*	6	,51941	5,2	83,1
V232n	1,00000	*	7	,50357	5,0	88,2
V233	1,00000	*	8	,42927	4,3	92,4
V234n	1,00000	*	9	,38983	3,9	96,3
V235	1,00000	*	10	,36545	3,7	100,0

Die Kommunalitäten sind beim Verfahren der Hauptkomponentenanalyse (Principal Components Analysis (PC)) zunächst alle gleich 1. Das bedeutet, daß die Varianz jedes Items bei der 10-faktoriellen Lösung vollständig erklärt wird.<sup>11</sup> Die Summe aller Eigenwerte ergibt die Gesamtvarianz aller Items, ist also in unserem Fall gleich 10. Die 10-faktorielle Lösung ist jedoch wenig aufschlußreich - wir hätten in diesem Fall auch gleich mit den einzelnen Items weiterrechnen können. In der Regel entscheidet man sich dafür, nur die Faktoren zu extrahieren, die mehr Varianz als die Varianz eines einzelnen Items erklären. Anders ausgedrückt - es werden die Faktoren extrahiert, deren Eigenwert (Eigenvalue) größer ist als 1. Dies ist das sog. Kaiser-Kriterium (siehe 3.1).

PC extracted 2 factors.

Auch in diesem Fall wurde das Kaiser-Kriterium angewendet - es werden 2 Faktoren extrahiert. Die folgende Faktorladungsmatrix (Factor Matrix) stellt die unrotierte Lösung dar (vgl. auch Kap. 5.4.2 im Skript):

Factor Matrix:

	Factor 1	Factor 2
V226n	,70535	-,22891
V227	,60937	,45461
V228n	,69906	-,29258
V229	,64812	,48154
V230n	,73863	-,27398
V231n	,68508	-,27524
V232n	,75937	-,23168
V233	,56269	,49564
V234n	,68935	-,35234
V235	,60335	,51003

In der unrotierten Lösung ist der erste Faktor immer auch der stärkste Faktor. Die Summe der quadrierten Faktorladungen entspricht dabei genau 4.525, also dem betragsgrößten

<sup>10</sup> Zu beachten ist hier die etwas unglückliche Anordnung des Outputs. Die Sternchen (\*) zwischen den Kommunalitäten und den Faktorladungen stellen eine Trennlinie dar - z.B. hat Faktor 1 in dieser Tabelle nichts mit V226 zu tun.

<sup>11</sup> Im Gegensatz dazu werden bei der Hauptachsenanalyse (Principal Axis Factoring (PAF)) zunächst die Kommunalitäten der Items geschätzt und während der Faktorenanalyse solange iterativ angepaßt, bis sie sich nicht mehr nennenswert ändern..

Eigenwert, wie er unter Initial Statistics ausgegeben wurde. Anders ausgedrückt: der erste Faktor erklärt mit 45,2% den größten Anteil der Gesamtvarianz aller Items.

Inhaltlich ist festzuhalten, daß alle Items relativ stark auf dem ersten Faktor laden. Der erste Faktor kann also als allgemeiner Selbstwert interpretiert werden. Die Vorzeichen der Ladungen auf Faktor 2 entsprechen genau der ursprünglichen Polung der Items. Damit ist schon hier ersichtlich, daß die positiv und negativ gepolten Items von den Befragten unterschiedlich beantwortet wurden.

Es folgt die Ausgabe der Final Statistics. In dieser Tabelle sind die Kommunalitäten der einzelnen Items aufgeführt, die sich aus der 2-faktoriellen Lösung ergeben:

Final Statistics:

Variable	Communality	*	Factor	Eigenvalue	Pct of Var	Cum Pct
V226n	, 54992	*	1	4, 52500	45, 2	45, 2
V227	, 57800	*	2	1, 41098	14, 1	59, 4
V228n	, 57429	*				
V229	, 65194	*				
V230n	, 62064	*				
V231n	, 54508	*				
V232n	, 63032	*				
V233	, 56228	*				
V234n	, 59934	*				
V235	, 62416	*				

Die beiden Faktoren erklären zusammen insgesamt 59,4% der Varianz der Items (Cum Pct). An den Kommunalitäten ist abzulesen, daß die Varianz jedes einzelnen Items zu über 50% durch die beiden Faktoren erklärt wird. Das ist ein akzeptables Ergebnis. Am „schlechtesten“ schneidet V231n („Mangelnde Selbstachtung“) mit 54,5% ab. Die meiste Varianz erklären die beiden Faktoren bei V229 („Bin eigentlich mit mir zufrieden“) mit 65,2%.

Da mehr als ein Faktor extrahiert wurde, wird zum Abschluß die rotierte Lösung (Rotated Factor Matrix) ausgegeben. Das VARIMAX-Rotationsverfahren maximiert die Ladungen eines Items auf jeweils einem Faktor und minimiert die Ladungen dieses Items auf den übrigen Faktoren. Dieses Verfahren liefert in der Regel die am einfachsten zu interpretierenden Lösungen:

VARI MAX rotation 1 for extraction 1 in analysis 1 - Kaiser Normalization.

Rotated Factor Matrix:

	Factor 1	Factor 2
V226n	, 70379	, 23366
V227	, 22139	, 72732
V228n	, 73646	, 17865
V229	, 23664	, 77197
V230n	, 75731	, 21708
V231n	, 71492	, 18433
V232n	, 74894	, 26344
V233	, 15947	, 73270
V234n	, 76406	, 12476
V235	, 18370	, 76839

Es fällt auf, daß die (im Sinne des allgemeinen Selbstwertes) negativ formulierten Items hoch auf Faktor 1, die positiv formulierten Items dagegen alle auf Faktor 2 laden. Hier gibt es zwei alternative Interpretationsmöglichkeiten:

1. *Negativer Selbstwert* ist inhaltlich etwas anderes als *das Fehlen eines positiven Selbstwertes*.
2. Die Zwei-Faktor-Lösung ist ein *Methodenartefakt*, d.h. die Befragten reagieren auf negativ formulierte Items anders als auf positive.

Welche der beiden Interpretationsmöglichkeiten die richtige ist, ist eine inhaltliche Frage, genauer eine Frage der Validität der Messung; sie ist nicht allein aufgrund der Faktorenanalyse zu entscheiden (siehe speziell hierzu die Ausführungen im Skript unter 5.4.2). Eine sinnvolle Strategie ist es in diesem Fall, zunächst zwei getrennte Skalen für positiven und negativen Selbstwert zu bilden und diese auf Kriteriumsvalidität zu untersuchen.

Als externe Kriterien können die in der gleichen Untersuchung erhobenen Skalen der *Kaufsucht* und der *materiellen Gütergebundenheit* dienen. Berechnet man die Korrelationen der beiden Selbstwertskalen<sup>12</sup> mit Kaufsucht und materieller Gütergebundenheit, so zeigt sich, daß negativer Selbstwert eine substantiell bedeutsame Korrelation mit Kaufsucht aufweist (.37); positiver Selbstwert korreliert dagegen nur schwach negativ mit Kaufsucht (-.12). In Bezug auf die materielle Gütergebundenheit ergibt sich für positiven Selbstwert eine signifikante, wenn auch relativ schwache Korrelation von .17, während negativer Selbstwert nicht signifikant mit materieller Gütergebundenheit korreliert (.05). Diese Ergebnisse sind eindeutige Hinweise darauf, daß es sich hier tatsächlich nicht um ein Methodenartefakt handelt, sondern daß negativer Selbstwert inhaltlich tatsächlich etwas anderes mißt, als fehlender positiver Selbstwert.

---

<sup>12</sup> Bei der Bildung der Subskalen wurden die Items in der ursprünglichen Polung verwendet.

#### 4. Beispiel: Parteiensympathie

In der ALLBUS-Studie 1994 wurden die Befragten unter anderem aufgefordert, ihre Sympathie zu den Parteien CDU (V144), SPD (V145), CSU (V146), F.D.P. (V147), „Bündnis 90 / Die Grünen“ (V148), „Die Republikaner“ (V149) und PDS (V150) auf einer 11-stufigen Ratingskala anzugeben.<sup>13</sup>

Als Arbeitshypothese stellen wir die Vermutung auf, daß es einen Faktor „politische Einstellung“ mit den Gegenpolen „linke politische Einstellung“ und „rechte politische Einstellung“ gibt, welcher der Parteiensympathie zugrunde liegt. Um die Hypothese der Eindimensionalität des Raumes der Parteiensympathie zu überprüfen, führen wir eine Faktorenanalyse (nach der Methode der Hauptkomponentenanalyse) mit den genannten sieben Items durch:

```
FACTOR
/VARIABLES v144 v145 v146 v147 v148 v149 v150 /MISSING LISTWISE /ANALYSIS
v144 v145 v146 v147 v148 v149 v150
/PRINT INITIAL CORRELATION KMO AIC EXTRACTION ROTATION
/FORMAT SORT BLANK(.20)
/PLOT EIGEN
/CRITERIA MINEIGEN(1) ITERATE(25)
/EXTRACTION PC
/CRITERIA ITERATE(25)
/ROTATION VARIMAX .
```

Neben den Angaben im letzten Beispiel haben wir mit diesem Befehl noch einen Factor Scree Plot (/PLOT EIGEN) angefordert. Wir erhalten folgenden Output:

```
----- F A C T O R   A N A L Y S I S -----
Analysis number 1   Listwise deletion of cases with missing values

Correlation Matrix:

          V144      V145      V146      V147      V148      V149      V150
V144      1,00000
V145      ,08357      1,00000
V146      ,81072      ,05565      1,00000
V147      ,48815      ,25643      ,48389      1,00000
V148      -,07312      ,34967      -,08815      ,21990      1,00000
V149      ,08673      -,09363      ,20790      ,06752      -,02710      1,00000
V150      -,16713      ,07845      -,15029      ,02673      ,28370      ,18856      1,00000
```

Beim ersten Blick auf die Korrelationsmatrix fällt auf, daß einige Items untereinander nur schwach korrelieren. An einigen Stellen tauchen nicht wie erwartet negative Korrelationen auf; so z.B. zwischen V144 (CDU) und V145 (SPD). Andererseits korreliert die Sympathie zur CSU (V146) tatsächlich negativ mit der Sympathie zur PDS (V150).

---

<sup>13</sup> Der genaue Wortlaut im Fragebogen: „Mich würde interessieren, wie Sie heute über die folgenden Parteien CDU, SPD, CSU, F.D.P., ‘Bündnis 90 / Die Grünen’, ‘Die Republikaner’ und PDS denken. Hier haben wir eine Art Thermometer. Je weiter Sie auf den weißen Kästchen nach oben gehen, umso sympathischer ist Ihnen die betreffende Partei; je weiter Sie auf den schwarzen Kästchen nach unten gehen, umso mehr lehnen Sie die Partei ab.“ Das vorgelegte „Thermometer“ hat Abstufungen von -5 bis +5, wobei +5 die höchste Sympathie darstellt.

Kaiser-Meyer-Olkin Measure of Sampling Adequacy = ,62098  
 Bartlett Test of Sphericity = 6438,7677, Significance = ,00000

Das KMO-Maß deutet darauf hin, daß diese Korrelationsmatrix für eine Faktorenanalyse nur „mäßig“ geeignet ist (vgl. Brosius 1995, S. 823). Der Bartlett-Test fällt trotz allem hochsignifikant aus, was bedeutet, daß diese Korrelationsmatrix hochsignifikant von der Einheitsmatrix abweicht.<sup>14</sup>

Anti-image Correlation Matrix:

	V144	V145	V146	V147	V148	V149	V150
V144	,61156						
V145	-,02458	,63308					
V146	-,73665	,01013	,59818				
V147	-,18977	-,16292	-,19526	,79645			
V148	,05139	-,29268	,07197	-,21833	,56876		
V149	,11765	,10227	-,24132	,00540	,03193	,42275	
V150	,05849	-,00998	,07750	-,05362	-,24234	-,23120	,58196

Measures of Sampling Adequacy (MSA) are printed on the diagonal.

Die MSA-Werte für die einzelnen Variablen sind ebenfalls eher mäßig. Besonders schlecht ist der MSA-Wert für V149 („Die Republikaner“). Bei einer weiteren Faktorenanalyse wäre es eventuell ratsam, auf dieses Item zu verzichten. Interessant ist hier die sehr hohe partielle Korrelation (.737) zwischen V144 (CDU) und V146 (CSU). Unabhängig von den Sympathien zum übrigen Parteienspektrum liegt hier offenbar eine starke gegenseitige Verbundenheit zwischen den Anhängern der beiden Parteien vor.

Extraction 1 for analysis 1, Principal Components Analysis (PC)

Initial Statistics:

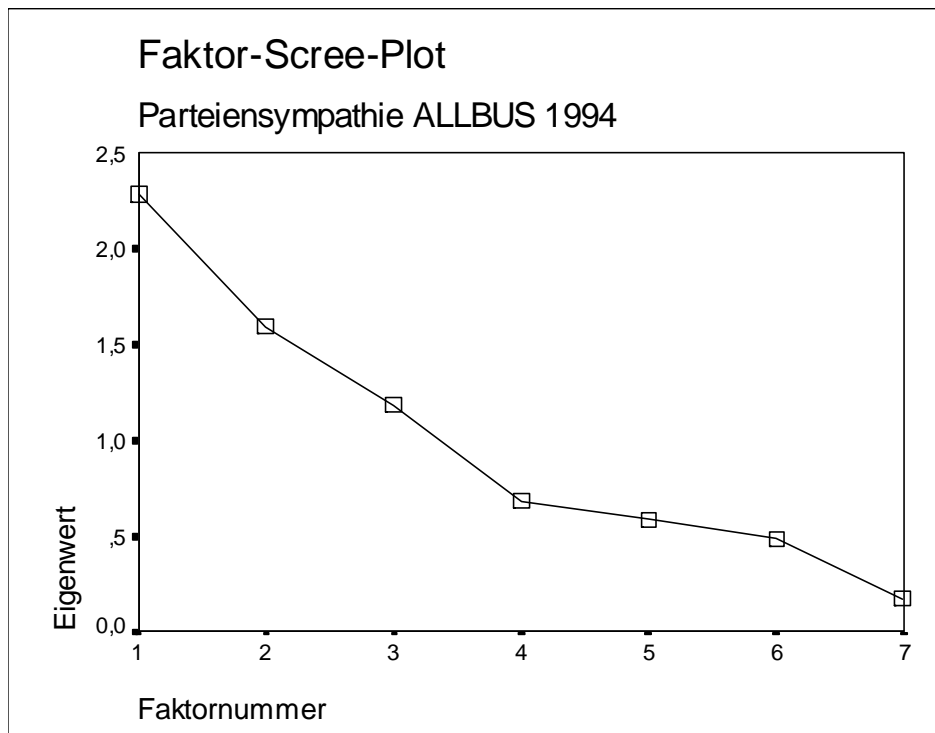
Variable	Communality	* Factor	Eigenvalue	Pct of Var	Cum Pct
V144	1,00000	* 1	2,28584	32,7	32,7
V145	1,00000	* 2	1,59799	22,8	55,5
V146	1,00000	* 3	1,18352	16,9	72,4
V147	1,00000	* 4	,68160	9,7	82,1
V148	1,00000	* 5	,58542	8,4	90,5
V149	1,00000	* 6	,48588	6,9	97,4
V150	1,00000	* 7	,17975	2,6	100,0

Die Initial Statistics zeigen an, daß es drei Faktoren mit Eigenwerten über 1 gibt. Es werden also drei Faktoren extrahiert, die gemeinsam 72,4% der Varianz der Items erklären. Daß dies eine adäquate Entscheidung ist, läßt sich sehr gut am folgenden Faktor-scree-plot erkennen:

---

<sup>14</sup> Der Bartlett-Test ist stark von der Größe der Stichprobe abhängig. Selbst eine Korrelationsmatrix mit äußerst geringen Korrelationen wäre bei der hier zugrundeliegenden Fallzahl von über 3000 Befragten noch signifikant von der Einheitsmatrix verschieden.

Hi - Res Chart # 1: Faktor-scree-plot<sup>15</sup>



Deutlich ist im Scree-Plot zu sehen, daß der Eigenwert der Faktoren insbesondere von Faktor 4 zu Faktor 3 zunimmt (man erkennt einen Knick über Faktornummer 4). Die Eigenwerte der Faktoren 4, 5, 6 und 7 sind dagegen gleichermaßen gering. Das Scree-Plot ist besonders gut geeignet, bei Faktorenanalysen mit einer sehr großen Zahl von Items die relevanten Faktoren zu erkennen.

Wir betrachten kurz die unrotierte Lösung der Faktorenanalyse:

PC extracted 3 factors.

Factor Matrix:

	Factor 1	Factor 2	Factor 3
V146	,90223		
V144	,89647		
V147	,74002	,31424	
V148		,82078	
V145	,23097	,65115	-,37229
V150		,58048	,56644
V149	,20499		,84141

Unter Optionen wurde angegeben, nur Werte mit einer Größe über .2 anzuzeigen (/BLANK (.20)). Sofern man sich alle Ladungen anzeigen läßt, entspricht auch hier die Summe der quadrierten Ladungen auf einem Faktor dem Eigenwert des Faktors. Die Kommunalitäten der Items ergeben sich aus der Zeilensumme der quadrierten Faktorladungen eines Items.

---

<sup>15</sup> Diese Grafik wird im Grafikkarussell von SPSS für Windows erzeugt und kann dort editiert werden.



Final Statistics:

Variable	Communality	*	Factor	Eigenvalue	Pct of Var	Cum Pct
V144	,83420	*	1	2,28584	32,7	32,7
V145	,61594	*	2	1,59799	22,8	55,5
V146	,85701	*	3	1,18352	16,9	72,4
V147	,65001	*				
V148	,67855	*				
V149	,75011	*				
V150	,68153	*				

Die erklärte Varianz der Items durch die drei Faktoren (Communality) ist auch in diesem Beispiel durchweg zufriedenstellend. Am „schlechtesten“ wird die Sympathie zur SPD (V145) durch diese Lösung erklärt (61,6%). Am „besten“ läßt sich die Sympathie zur CSU (V146) aus den drei extrahierten Faktoren ableiten (85,7%). Zur Interpretation der Faktoren betrachten wir nun wieder die rotierte Faktormatrix (zur besseren Übersicht haben wir die Labels der Items in den Output eingefügt):

VARIMAX rotation 1 for extraction 1 in analysis 1 - Kaiser Normalization.

Rotated Factor Matrix:

		Factor 1	Factor 2	Factor 3
CSU	V146	,91992		
CDU	V144	,91063		
F. D. P.	V147	,69218	,41006	
B90/GRÜNE	V148		,80049	
SPD	V145		,75489	
REP	V149	,21437	-,22722	,80779
PDS	V150	-,22154	,36732	,70536

Factor 1 läßt sich als „rechts-konservative Einstellung“ deuten. Factor 2 kann man als „links-liberale Einstellung“ interpretieren. Interessant ist, daß sich die Sympathie für die F.D.P. zwar überwiegend durch den konservativen Faktor erklären läßt ( $.69^2 = 47,6\%$ ), daneben aber auch (noch? [1994!]) ein links-liberales Element in der Sympathie für die F.D.P. zu erkennen ist ( $.41^2 = 16,8\%$ ). Factor 3 spiegelt offenbar eine „Nähe zu extremistischen Parteien“ wider. Beide „Protestparteien“, die Republikaner und die PDS, laden hoch auf diesem Faktor. V149 (Die Republikaner) hat erwartungsgemäß zusätzlich eine (schwach) negative Ladung auf Factor 2 und eine (schwach) positive Ladung auf Factor 1 - entsprechend Umgekehrtes gilt für V150 (PDS).

Unsere Arbeitshypothese, die Sympathie zu den Parteien ließe sich durch einen bipolaren „Links-Rechts“-Faktor hinreichend erklären, ist offenbar widerlegt worden. Sympathien zum rechts-konservativen oder links-liberalen Parteienspektrum schließen sich nicht gegenseitig aus, sondern stellen zwei unabhängige Faktoren dar. Darüber hinaus gibt es eine weitere unabhängige Dimension, die als politischer Protest, bzw. Extremismus gedeutet werden kann. Der Parteien-Raum der BRD ist also dreidimensional.