

Kapitel 26

Faktorenanalyse

In den Wirtschafts- und Sozialwissenschaften hat man es häufig mit Begriffen und Sachverhalten zu tun, die einen hohen Grad an Komplexität aufweisen und sich daher nicht durch eine einzelne Variable, deren Werte zum Beispiel durch eine Frage in einem Fragebogen oder durch eine andere einfache Messung ermittelt wurden, abbilden lassen. Begriffe wie „Kreativität“, „Intelligenz“ und „Qualifikation“ oder auch scheinbar einfachere Begriffe wie „Konjunktur“ oder „Wohlfahrt“ repräsentierten für den Experten jeweils eine umfangreiche Theorie mit einem entsprechend großen Bündel von Variablen, die zur adäquaten Darstellung dieser Begriffe zu berücksichtigen sind. Um derart komplexe Begriffe empirisch zu untersuchen, müssen sie daher zumeist in eine Vielzahl einzelner Variablen zerlegt werden, da sie durch eine einzige nicht angemessen zu erfassen sind.

Die Faktorenanalyse folgt diesen Überlegungen, geht aber gerade den umgekehrten Weg: Der Ausgangspunkt einer Faktorenanalyse ist eine große Anzahl von Variablen, von denen a priori nicht bekannt ist, ob und in welcher Weise sie miteinander zusammenhängen. Mit der Faktorenanalyse wird untersucht, ob sich unter den betrachteten Variablen Gruppen von Variablen befinden, denen jeweils eine komplexe Hintergrundvariable wie zum Beispiel „Kreativität“ oder „Wohlfahrt“ zugrunde liegt. Solche Hintergrundvariablen werden im Rahmen der Faktorenanalyse als *Faktoren* bezeichnet. Das Ziel einer Faktorenanalyse ist es, den hohen Grad an Komplexität, der durch eine Vielzahl von Variablen abgebildet wird, dadurch handhabbar und oft auch erst interpretierbar zu machen, daß die Variablen auf möglichst wenige Faktoren, die hinter ihnen stehen, reduziert werden.

26.1 Die Beispieldaten: Frühgeburtenstudie

An der Kinderklinik der Hamburger Universitätsklinik Eppendorf (UKE) wird unter der Leitung von *Prof. Dr. F. J. Schulte* eine Untersuchung an Frühgeborenen durchgeführt. Dabei werden Frühgeborene untersucht, die in der Zeit von Juli 1983 bis Juni 1986 geboren wurden. Das kennzeichnende Merkmal einer Frühgeburt ist nicht etwa die Länge der Gestationszeit, sondern das Gewicht des Kindes bei der Geburt. Säuglinge mit einem Geburtsgewicht ≤ 1.500 g gelten als

frühgeboren. Für diese Kinder wurden zum Geburtszeitpunkt verschiedene Daten erhoben. Darüber hinaus wurden und werden die Kinder in verschiedenen Jahresabständen in neurologischer und psychologischer Hinsicht beobachtet und untersucht. Es handelt sich damit um eine Längsschnittuntersuchung. Die Aufklärung des Zusammenhangs zwischen frühgeburtlichen Daten einerseits und neurologischen sowie psychosozialen Faktoren andererseits hat wegen der medizinischen Entwicklung der vergangenen Jahrzehnte an Bedeutung gewonnen, denn während früher nur ca. 15% der Frühgeborenen überlebten, liegt der Anteil inzwischen bei ca. 65%. Die vor allem auf verstärkten Einsatz von Intensivmedizin zurückzuführende Entwicklung hat allerdings vermutlich den Preis, daß damit der Anteil neurologischer und psychosozialer Störungen in der Kindesentwicklung zugenommen hat.²⁸⁶

Neben anderen Fragestellungen interessierte die mit der Untersuchung befaßte Forschergruppe, ob sich zwischen bestimmten perinatalen Daten und der späteren Intelligenzentwicklung ein Zusammenhang erkennen läßt. Zu diesem Zweck wurden mit allen in die Untersuchung einbezogenen Frühgeborenen in ihrem sechsten Lebensjahr verschiedene Tests durchgeführt, deren Ergebnisse in den in Abbildung 26.1 aufgeführten Variablen festgehalten sind.

Variablen	Bedeutung
cmmt	Columbia Mental Maturity Test
fewtw	Frostigs Entwicklungs-, Visualisierungs- und Wahrnehmungstest
moe	Moering-Test
petwetw	Psychologischer Entwicklungstest: Wörter ergänzen
petgtw	Psychologischer Entwicklungstest: Grammatik
petzfgtw	Psychologischer Entwicklungstest: Zahlenfolgedächtnis
petoftw	Psychologischer Entwicklungstest: Objekte finden
petsfgtw	Psychologischer Entwicklungstest: Symbolfolgedächtnis
frkonz	Frankfurter Konzentrationstest für 5jährige
mannztw	Mannzeichnen
prozsum	Bildererkennung in einem zu einem bestimmten Prozentsatz unvollständigen Mosaik; Summe aus verschiedenen Einzeltests

Abbildung 26.1: Bedeutung ausgewählter Variablen aus der Datei „Frühgeborene.sav“



Diese und weitere Variablen der Frühgeburtenstudie finden Sie auf der beigelegten CD in der Datendatei *Frühgeborene.sav*. In diesem Zusammenhang ist der o.a. Forschungsgruppe und insbesondere Herrn *Dr. Dieter Lutz* vom Psychologischen Institut der Universität Hamburg zu danken, daß sie die Daten nicht nur für Beispiele in diesem Buch, sondern auch zur Weitergabe an dessen Leser zur Verfügung gestellt und damit einen Beitrag zur praxisnahen Statistikausbildung geleistet haben!

²⁸⁶ Wesentliche Teilergebnisse dieser von der DFG geförderten Untersuchung finden sich bei *Veelken, Norbert* (1992): Entwicklungsprognose von Kindern mit einem Geburtsgewicht unter 1501 g. Eine regional repräsentative Follow-Up-Studie über 371 Kinder. Habilitationsschrift am Fachbereich Medizin der Universität Hamburg.

26.2 Das Modell der Faktorenanalyse

Im allgemeinen sind weder die Art noch die Anzahl der Faktoren, die mit Hilfe der Faktorenanalyse bestimmt werden sollen, im voraus bekannt. Aus didaktischen Gründen werden hier jedoch bereits an dieser Stelle die Faktoren mitgeteilt, deren (mögliche) Existenz erst als Ergebnis der Faktorenanalyse ermittelt werden soll. Dadurch kann das Modell der Faktorenanalyse im folgenden leichter dargestellt und nachvollzogen werden. Hinter den in Abbildung 26.1 aufgeführten Variablen stehen (möglicherweise) die beiden folgenden Faktoren:

- Allgemeine Intelligenz (AI)
- Sprachliche Intelligenz (SI)

Wenn diese beiden Faktoren hinter den elf Variablen aus Abbildung 26.1 stehen, bedeutet dies, daß sie die Variablen bestimmen und im wissenschaftlichen Sinne erklären. Damit läßt sich der Zusammenhang zwischen den Faktoren und den Variablen - ähnlich wie bei einer Regressionsanalyse - durch ein System von Gleichungen beschreiben. Wenn jede der elf Stichprobenvariablen durch die beiden Faktoren erklärt wird, so läßt sich für jede Variable eine Gleichung formulieren, die diesen Zusammenhang beschreibt. Für die erste Variable aus Abbildung 26.1 könnte die Gleichung z.B. folgendermaßen lauten:

$$cmmt = a_1 \cdot AI + a_2 \cdot SI + u_{cmmt}$$

Es fällt unmittelbar die Ähnlichkeit mit einer Regressionsgleichung auf, die auch inhaltlich durch folgenden Zusammenhang zum Ausdruck kommt: Die Faktoren *AI* und *SI* sind wie erklärende Variablen (Prädiktoren) zu betrachten, durch die die zu erklärende (abhängige) Variable *cmmt* erklärt wird. Die Koeffizienten a_1 und a_2 entsprechen den Regressionskoeffizienten einer Regressionsgleichung, und der dritte Faktor, u_{cmmt} , entspricht den Residuen (oder den Fehlern) einer Regressions-schätzung. Dieser Fehlerterm muß deshalb in die Gleichung aufgenommen werden, weil im allgemeinen nicht zu erwarten ist, daß die zu erklärende Variable (in diesem Fall *cmmt*) vollständig durch die übrigen Faktoren (in diesem Fall also *AI* und *SI*) erklärt wird. In der Faktorenanalyse wird der Fehlerterm (u_{cmmt}), der sozusagen den durch die übrigen Faktoren nicht erklärten Rest darstellt, als *Einzelrestfaktor* (unique factor) bezeichnet, weil er allein in der Gleichung für die Variable *cmmt* enthalten ist.²⁸⁷ In Gleichungen, durch die empirische Ergebnisse wiedergegeben werden, wird der Restfaktor im allgemeinen nicht mit aufgeführt.

Die beiden anderen Faktoren *AI* und *SI* werden als *gemeinsame Faktoren* (common factors) bezeichnet, da sie zur Erklärung jeder im Modell enthaltenen Variablen herangezogen werden. Dies schließt freilich nicht aus, daß einzelne Faktoren für verschiedene Variablen nur einen sehr geringen Erklärungsgehalt und damit nur einen sehr geringen Einfluß haben. Ein geringer Einfluß eines Faktors kommt in einem entsprechend geringen Wert des Faktorkoeffizienten (a_1 bzw. a_2) zum Ausdruck.

²⁸⁷ Zwar enthalten auch die Gleichungen der übrigen Variablen einen Fehlerterm, jedoch ist dies nicht der Restfaktor u_{cmmt} , sondern u_{fewtw} (für die zweite Variable aus Abbildung 26.1) etc.

Wie bereits gesagt, sind die Faktoren (in diesem Beispiel also *AI* und *SI*) nicht im voraus bekannt, sondern sollen gerade erst durch die Faktorenanalyse ermittelt werden. Der bei der Berechnung von Faktoren zugrundeliegende Gedanke sei im folgenden kurz skizziert, ohne daß das - zum Teil recht komplizierte - mathematische Verfahren exakt dargestellt wird. Zunächst werden lineare Kombinationen der Beobachtungsvariablen gebildet. Für Variablen, die miteinander hohe Korrelationskoeffizienten aufweisen, wird angenommen, daß sie einen gemeinsamen Faktor haben. Umgekehrt wird für Variablen, die nur schwach miteinander korrelieren, unterstellt, daß sie keinen Faktor miteinander gemein haben. Für die konkrete Ermittlung der Faktoren stehen verschiedene Schätzverfahren zur Verfügung, die weiter unten behandelt werden.²⁸⁸ Im Prinzip laufen jedoch alle Verfahren letztlich darauf hinaus, die Koeffizienten c_i der folgenden Gleichung, die den Zusammenhang für den ersten Faktor *AI* beschreibt, zu berechnen.

$$AI = c_1 \cdot \text{cmmt} + c_2 \cdot \text{fewtw} + \dots + c_{11} \cdot \text{prozsum}$$

Nach dieser Gleichung ist es formal möglich, daß alle elf Variablen der Stichprobe zur Erklärung des Faktors *AI* (Allgemeine Intelligenz) beitragen. Das Ziel und die Hoffnung der Faktorenanalyse besteht allerdings darin, daß der Faktor nur von einem Teil der Variablen bestimmt wird. Umgekehrt ist eine erfolgreiche Faktorenanalyse dadurch gekennzeichnet, daß die Vielzahl der in der Stichprobe relevanten Variablen durch nur wenige Faktoren repräsentiert werden. Es wäre durch die Faktorenanalyse nichts gewonnen, wenn im vorliegenden Beispiel elf Faktoren benötigt würden, um die Zusammenhänge zu charakterisieren, denn dann könnte man auch direkt mit den elf Variablen arbeiten. Ziel war es ja schließlich, den Grad an Komplexität dadurch zu verringern, daß anstatt der Vielzahl von Einzelvariablen nur eine begrenzte Zahl von Hintergrundvariablen (Faktoren) betrachtet werden muß.

Weiterhin kann eine Faktorenanalyse nur dann als erfolgreich angesehen werden, wenn die dadurch ermittelten Faktoren inhaltlich sinnvoll interpretierbar sind. Dies stellt eines der schwierigsten Probleme bei der Durchführung einer Faktorenanalyse dar. Diese Problematik wurde bisher dadurch verdeckt, daß bereits zu Beginn zwei Faktoren mit den sprechenden Namen *Allgemeine Intelligenz* und *Sprachliche Intelligenz* eingeführt wurden. Die Faktorenanalyse liefert dagegen zunächst nur Faktoren, die von SPSS als *Faktor 1*, *Faktor 2* etc. bezeichnet werden. Es ist die Aufgabe des Analytikers, diese Faktoren inhaltlich sinnvoll zu interpretieren und ihnen dann gegebenenfalls aussagekräftige Namen zu geben.

26.3 Die vier Schritte einer Faktorenanalyse

Üblicherweise wird die Faktorenanalyse in vier Schritten durchgeführt, an denen sich auch die nachfolgende Darstellungsweise orientiert. Dies bedeutet jedoch nicht, daß die Faktorenanalyse ausschließlich in dieser Weise durchgeführt werden

²⁸⁸ Siehe Abschnitt 26.5, *Faktorextraktion*, S. 647, insbesondere auch den Unterpunkt *Verschiedene Methoden der Faktorextraktion*, S. 654.

kann. Vielmehr ist es mit der SPSS-Prozedur FAKTORENANALYSE ebenso möglich, selbst eine sehr komplexe Faktorenanalyse in einem einzigen Durchlauf auszuführen. Andererseits ermöglichen es die zahlreichen bei der Prozedur FAKTORENANALYSE zur Verfügung stehenden Optionen, die einzelnen Schritte sehr genau auszuwerten. Die vier üblichen Schritte einer Faktorenanalyse sind die folgenden:

- **Korrelationsmatrizen:** Sie können die Korrelationsmatrizen für alle in die Faktorenanalyse einbezogenen Variablen berechnen und ausgeben lassen. An den Korrelationsmatrizen kann abgelesen werden, welche Variablen möglicherweise unberücksichtigt bleiben sollten, weil sie mit den übrigen Variablen nur sehr geringe Korrelationen aufweisen (siehe Abschnitt 26.4, *Korrelationsmatrizen*, S. 644).
- **Faktorextraktion:** Dieser Schritt wird im allgemeinen als „Ziehen“ oder „Extrahieren“ von Faktoren bezeichnet. Da es verschiedene Methoden der Faktorextraktion gibt, müssen Sie bei diesem Schritt in den Dialogfeldern der Prozedur angeben, welche Extraktionsmethode zur Anwendung kommen soll. Verschiedene statistische Maßzahlen, die Sie sich bei diesem Schritt ausgeben lassen können, zeigen dann auch auf, ob das angenommene Faktorenmodell geeignet ist, die Variablen auf einfache Weise zu repräsentieren. Die Faktorextraktion wird in Abschnitt 26.5, *Faktorextraktion*, S. 647 dargestellt.
- **Rotation:** Die im zweiten Schritt gefundenen Faktoren sind häufig zunächst schwierig zu interpretieren. Um die Interpretation zu erleichtern, macht man sich den Umstand zunutze, daß die Faktoren Kunstgebilde sind, die sich verzerrungsfrei so transformieren lassen, daß sie in unterschiedlichen Koordinatensystemen dargestellt werden können. Durch eine geeignete Transformation gelingt es häufig, die *Verbindung* zu den Beobachtungsvariablen deutlicher aufzuzeigen und damit die Interpretation der Faktoren zu erleichtern. Dieser Schritt wird als Rotation bezeichnet, da hierbei die Koordinatenachsen in gewisser Weise gedreht werden. Siehe hierzu im einzelnen Abschnitt 26.6, *Rotation*, S. 655.
- **Faktorwerte:** Obwohl die Faktoren in gewisser Weise als komplexe, die beobachteten Variablen prägende *Hintergrundvariablen* aufgefaßt werden können, läßt sich das wesentliche Ziel einer Faktorenanalyse im Prinzip erreichen, ohne jemals konkrete Werte dieser besonderen Kunstvariablen zu ermitteln und darzustellen. Andererseits besteht das Ziel einer Faktorenanalyse oftmals darin, die ermittelten Faktoren anschließend in anderen statistischen Verfahren zu verwenden, um beispielsweise weitere, in die Faktorenanalyse nicht einbezogene Variablen zu erklären oder umgekehrt die ermittelten Faktoren durch andere Variablen erklären zu lassen. Für diese Zwecke können Sie konkrete Faktorwerte berechnen und ggf. als Variable in der Datendatei speichern. Siehe hierzu im einzelnen Abschnitt 26.7, *Schätzung von Faktorwerten*, S. 661.

26.4 Korrelationsmatrizen

Korrelationskoeffizienten

Abbildung 26.2 gibt die Korrelationsmatrix für die elf Variablen aus der Stichprobe wieder. Um diese Matrix sowie die anderen in diesem Abschnitt betrachteten Ergebnisse zu erstellen, nehmen Sie folgende Einstellungen vor:

- **Prozedur aufrufen:** Um eine Faktorenanalyse durchzuführen, wählen Sie den Befehl



STATISTIK
DIMENSIONSREDUKTION ▶
FAKTORENANALYSE...

- **Variablen:** Verschieben Sie die elf in Abbildung 26.1, S. 640 aufgeführten Variablen in die Liste *Variablen*.
- **Deskriptive Statistiken:** Wählen Sie in dem Dialogfeld der Schaltfläche *Deskriptive Statistik* in der Gruppe *Korrelationsmatrix* die Optionen *Koeffizienten*, *KMO* und *Bartlett-Test auf Sphärizität* und *Anti-Image*.

Korrelationsmatrix

	CMMT	FEWTW	FRKONZ	MANNZTW	MOE	PETGTTW	PETOFTW	PETSFGTW	PETWETW	PETZFGTW	PROZSUM
Korrelation	1,000	,499	,428	,495	,413	,524	,515	,507	,426	,330	,246
CMMT		1,000	,525	,505	,426	,406	,582	,421	,275	,173	,296
FEWTW			1,000	,486	,325	,336	,581	,480	,236	,136	,269
FRKONZ				1,000	,300	,339	,587	,447	,191	,175	,358
MANNZTW					1,000	,370	,418	,360	,419	,257	,229
MOE						1,000	,400	,473	,641	,422	,268
PETGTTW							1,000	,573	,320	,229	,451
PETOFTW								1,000	,362	,260	,329
PETSFGTW									1,000	,438	,245
PETWETW										1,000	,105
PETZFGTW											1,000
PROZSUM											

Abbildung 26.2: Matrix der Korrelationskoeffizienten für die elf Stichprobenvariablen

Diese Korrelationsmatrix gibt einen ersten Überblick darüber, welche Variablen stark und welche nur schwach miteinander korrelieren.²⁸⁹ So ist etwa zu erkennen, daß zwischen den Variablen *cmmt* und *fewtw* eine relative starke Korrelation besteht. Zusätzlich weisen beide Variablen eine relativ hohe Korrelation zur Variablen *frkonz* auf. Dagegen ist jede der drei Variablen nur deutlich schwächer mit der Variablen *prozsum* korreliert. Würden alle ausgewiesenen Korrelationskoeffizienten nur sehr geringe absolute Werte aufweisen, wäre es wenig sinnvoll, die Faktorenanalyse fortzusetzen, da gemeinsame Faktoren nur für solche Variablen existieren, die relativ stark miteinander korreliert sind.

²⁸⁹ Zum Begriff der Korrelation siehe Kapitel 21, *Bivariate Korrelationen*, S. 497.

Test auf Sphärizität

In der Korrelationsmatrix wurden für zahlreiche Variablenpaare relativ starke Korrelationen beobachtet. Dennoch ist es möglich, daß sich die ausgewiesenen Korrelationen nur zufällig in der zugrundeliegenden Stichprobe ergeben haben, obwohl in der Grundgesamtheit keinerlei Zusammenhang zwischen den Variablen besteht und sämtliche Korrelationskoeffizienten einen Wert von null haben. Mit dem *Bartlett-Test auf Sphärizität* kann die Hypothese getestet werden, nach der alle Korrelationskoeffizienten zwischen den Variablen in der Grundgesamtheit den Wert 0 haben. Das Ergebnis dieses Tests wird in Abbildung 26.3 wiedergegeben.²⁹⁰

Maß der Stichprobeneignung nach Kaiser-Meyer-Olkin.		,897
Bartlett-Test auf Sphärizität	Ungefähres Chi-Quadrat	1281,153
	df	55
	Signifikanz nach Bartlett	,000

Abbildung 26.3: KMO-Maß und Bartlett-Test auf Sphärizität

Die Testgröße von Bartletts Test ist ein Chi-Quadrat-Wert, der mit 1281 außerordentlich hoch ist. Dementsprechend wird ein Signifikanzwert von 0,000 ausgewiesen. Dies ist so zu interpretieren, daß die Hypothese, alle Korrelationen zwischen den elf Variablen seien in der Grundgesamtheit gleich 0, mit einer Irrtumswahrscheinlichkeit von 0,000 zurückgewiesen werden kann. Umgekehrt kann man also davon ausgehen, daß zumindest zwischen einigen der elf Variablen auch in der Grundgesamtheit Korrelationen bestehen.

Anti-Image-Korrelationsmatrix

In Abbildung 26.4 ist die sogenannte Anti-Image-Korrelationsmatrix wiedergegeben.²⁹¹ Dieser Matrix liegen die partiellen Korrelationskoeffizienten zugrunde. Der partielle Korrelationskoeffizient eines Variablenpaares gibt die Korrelation zwischen den beiden betreffenden Variablen an, die sich ergibt, wenn die linearen Einflüsse der übrigen Variablen zuvor ausgeschaltet wurden. Dies bedeutet, daß die nicht erklärten Teile der Variablen miteinander korreliert werden. Dies wiederum sind die Einzelrestfaktoren, von denen bereits in Abschnitt 26.2, *Das Modell der Faktorenanalyse*, S. 641 die Rede war. Für den Fall, daß zwei Variablen einen gemeinsamen Faktor aufweisen, müßte ihr partieller Korrelationskoeffizient annähernd null sein, weil die gemeinsame Streuung bei der Berechnung des partiellen Korrelationskoeffizienten zuvor eliminiert wird und nur die Einzelrestfakto-

²⁹⁰ Das ebenfalls in Abbildung 26.3 aufgeführte KMO-Maß wird weiter unten in diesem Abschnitt betrachtet.

²⁹¹ In der von SPSS erstellten Form umfaßt die Anti-Image-Korrelationsmatrix die beiden Teile *Anti-Image Covariance* und *Anti-Image Correlation*, der erste dieser beiden Teile wurde für die Darstellung in Abbildung 26.4 jedoch ausgeblendet.

ren übrigbleiben. Diese wiederum - das ergibt sich aus ihrer Definition - dürfen miteinander keine nennenswerten Korrelationen aufweisen.

Anti-Image-Matrizen

Anti-Image-Korrelation

	CMMT	FEWTW	FRKONZ	MANNZTW	MOE	PETGTTW	PETOFTW	PETSFGTW	PETWETW	PETZFGTW	PROZSUM
CMMT	,930 ^a	-,141	-,036	-,201	-,096	-,168	-,079	-,149	-,091	-,094	,072
FEWTW	-,141	,911 ^a	-,203	-,140	-,181	-,125	-,209	,036	,055	,051	-,011
FRKONZ	-,036	-,203	,918 ^a	-,131	-,020	-,022	-,235	-,158	,004	,059	,039
MANNZTW	-,201	-,140	-,131	,909 ^a	,012	-,047	-,225	-,058	,130	-,021	-,134
MOE	-,096	-,181	-,020	,012	,925 ^a	,026	-,098	-,044	-,218	-,046	-,009
PETGTTW	-,168	-,125	-,022	-,047	,026	,857 ^a	,040	-,161	-,457	-,154	-,037
PETOFTW	-,079	-,209	-,235	-,225	-,098	,040	,892 ^a	-,229	-,023	-,042	-,242
PETSFGTW	-,149	,036	-,158	-,058	-,044	-,161	-,229	,931 ^a	-,019	-,028	-,064
PETWETW	-,091	,055	,004	,130	-,218	-,457	-,023	-,019	,808 ^a	-,210	-,096
PETZFGTW	-,094	,051	,059	-,021	-,046	-,154	-,042	-,028	-,210	,892 ^a	,059
PROZSUM	,072	-,011	,039	-,134	-,009	-,037	-,242	-,064	-,096	,059	,894 ^a

a. Maß der Stichprobeneignung

Abbildung 26.4: Anti-Image-Korrelationsmatrix

In der Anti-Image-Korrelationsmatrix werden nicht die partiellen Korrelationskoeffizienten selbst, sondern ihre negativen Werte ausgewiesen. Wenn das Faktorenmodell geeignet ist, sollten diese Werte nahe bei null liegen. Ein zusammenfassendes Maß hierfür ist das Kaiser-Mayer-Olkin-Maß (KMO), das folgendermaßen berechnet wird:

$$KMO = \frac{\sum \sum r_{ij}^2}{\sum \sum r_{ij}^2 + \sum \sum a_{ij}^2} \quad \text{für } i \neq j$$

Dabei bezeichnet r_{ij} den einfachen Korrelationskoeffizienten zwischen den Variablen i und j , und a_{ij} repräsentiert die partiellen Korrelationskoeffizienten. Weil Korrelationen einer Variablen mit sich selbst (die stets den Wert 1 haben) natürlich nicht berücksichtigt werden dürfen, werden die Kombinationen mit $i = j$ aus der Summenbildung ausgeschlossen.

Das KMO-Maß kann höchstens den Wert 1 annehmen. Ein Wert in der Nähe von 1 wird dann erreicht, wenn die partiellen Korrelationskoeffizienten sehr klein sind. Umgekehrt nimmt das KMO-Maß bei großen partiellen Korrelationskoeffizienten einen kleinen Wert an. Ein kleiner KMO-Wert zeigt damit an, daß die Variablenauswahl für eine Faktorenanalyse nicht gut geeignet ist. Kaiser selbst hat die in Abbildung 26.5 wiedergegebene Beurteilung für die Ergebnisse des KMO-Maßes vorgelegt.²⁹²

Das KMO-Maß für die elf in diesem Beispiel betrachteten Variablen wurde bereits in Abbildung 26.3 ausgewiesen. Legt man die Beurteilung von Kaiser zugrunde, ist der Wert mit 0,897 *recht gut*, so daß die Zusammenstellung der Variablen für ein faktoranalytisches Modell durchaus geeignet zu sein scheint.

²⁹² Vgl. Kaiser, H.F. (1974): An Index of Factorial Simplicity, in: Psychometrika, Vol. 39, S. 31-36.

Wert	Beurteilung
0,9 bis 1,0	fabelhaft (marvelous)
0,8 bis unter 0,9	recht gut (meritorious)
0,7 bis unter 0,8	mittelprachtig (middling)
0,6 bis unter 0,7	maig (mediocre)
0,5 bis unter 0,6	schlecht (miserable)
unter 0,9	inakzeptabel (unacceptable)

Abbildung 26.5: Beurteilung der Ergebnisse des KMO-Maes nach Kaiser

Bevor Sie das gewahlte Modell jedoch endgultig akzeptieren, sollten Sie noch die MSA-Werte betrachten, die in der Hauptdiagonalen der Anti-Image-Korrelationsmatrix aus Abbildung 26.4 ausgewiesen werden (siehe auch die Funote unterhalb der Matrix). MSA ist die Abkurzung fur *Measure of Sampling Adequacy* (Ma fur die Angemessenheit der Stichprobe). Die MSA-Werte werden im Prinzip genauso berechnet, wie das gerade beschriebene KMO-Ma, mit dem Unterschied, da es sich lediglich auf jeweils eine Variable statt auf alle Variablen insgesamt bezieht:

$$MSA_i = \frac{\sum r_{ij}^2}{\sum r_{ij}^2 + \sum a_{ij}^2} \quad \text{fur } i \neq j$$

Es werden also so viele MSA-Werte berechnet, wie es Variablen im Faktorenmodell gibt. Der Matrix in Abbildung 26.4 ist zum Beispiel zu entnehmen, da fur die Variable *moe* der MSA-Wert 0,925 gilt, der nach der Beurteilung von Kaiser (die entsprechend fur MSA-Werte Gultigkeit hat) als *fabelhaft* anzusehen ist. Der kleinste in der Matrix ausgewiesene MSA-Wert betragt 0,808 (Variable *petwetw*) und ist damit immer noch *recht gut*. Die MSA-Werte bieten daher im vorliegenden Beispiel keinen Anla, eine oder mehrere Variablen aus dem faktoranalytischen Modell auszuschlieen.

26.5 Faktorextraktion

Zur Berechnung der Faktoren einer Faktorenanalyse, die zumeist als Faktorextraktion bezeichnet wird, werden in der Literatur verschiedene Verfahren vorgeschlagen. Jedes dieser Verfahren hat seine Vor- und Nachteile. Die wichtigsten dieser Verfahren werden weiter unten in diesem Abschnitt kurz skizziert. SPSS benutzt per Voreinstellung das Verfahren der Hauptkomponentenanalyse (Principal Component Analysis), auf das die Demonstration des Beispiels beschrankt bleiben wird. Dies soll jedoch nicht zum Ausdruck bringen, da das Verfahren der Hauptkomponentenanalyse den ubrigen Verfahren uberlegen ist.

Bei dem Verfahren der Hauptkomponentenanalyse werden lineare Kombinationen der Variablen gebildet. Als erste Hauptkomponente (= Faktor) wird diejenige ausgewiesen, die den groten Teil der Gesamtstreuung aller Variablen im statistischen Sinne erklart. Die zweite Hauptkomponente ist entsprechend diejenige, die den zweitgroten Teil erklart etc. Formal konnen so viele Hauptkomponenten und damit Faktoren berechnet werden, wie in dem Faktorenmodell Variablen enthalten

sind. In der tabellarischen Übersicht *Erklärte Gesamtvarianz*, die sich zur Hälfte auf den Anfangsstatus der Faktorextraktion bezieht, werden auch tatsächlich Maßzahlen für so viele Hauptkomponenten ausgegeben, wie der Faktorenanalyse Variablen zugrundeliegen. Um diese Übersicht, die in Abbildung 26.6 wiedergegeben wird, und die anderen in diesem Abschnitt betrachteten Ergebnisse zu erhalten, nehmen Sie in den Dialogfeldern die Faktorenanalyse folgende Einstellungen vor:

- **Deskriptive Statistiken:** Kreuzen Sie in dem Dialogfeld der Schaltfläche *Deskriptive Statistik* die Option *Anfangslösung* an.
- **Extraktion:** Wählen Sie in dem Dialogfeld *Extraktion* die Optionen *Nicht rotierte Faktorenlösung* und *Screepplot*.

Eigenwerte und Kommunalitäten

Die linke Hälfte der Tabelle in Abbildung 26.6 (*Anfängliche Eigenwerte*) bezieht sich auf ein Anfangsstadium der Faktorextraktion, in dem so viele Faktoren wie Variablen vorhanden sind.

Komponente	Anfängliche Eigenwerte			Summen von quadrierten Faktorladungen für Extraktion		
	Gesamt	% der Varianz	Kumulierte %	Gesamt	% der Varianz	Kumulierte %
1	4,879	44,355	44,355	4,879	44,355	44,355
2	1,389	12,624	56,979	1,389	12,624	56,979
3	,824	7,489	64,468			
4	,721	6,553	71,021			
5	,620	5,636	76,658			
6	,558	5,070	81,727			
7	,529	4,805	86,532			
8	,433	3,936	90,468			
9	,404	3,673	94,141			
10	,345	3,139	97,279			
11	,299	2,721	100,000			

Extraktionsmethode: Hauptkomponentenanalyse.

Abbildung 26.6: Eigenwerte der Faktoren für Lösungen mit elf und zwei Faktoren

Der *Eigenwert* (*Eigenvalue*) eines Faktors gibt an, welcher Betrag der Gesamtstreuung aller Variablen des Faktorenmodells durch diesen Faktor erklärt wird. Hierbei ist zu beachten, daß zu diesem Zweck - wie im übrigen für die gesamte Berechnung der Faktorenanalyse - die Variablen zuvor einer Transformation in sogenannte Z-Werte unterworfen werden. Dies bedeutet im Ergebnis, daß jede Variable anschließend eine Standardabweichung von 1 und einen Mittelwert von 0 hat. Damit beträgt die gesamte zu erklärende Streuung in dem vorliegenden Beispiel mit elf Variablen ebenfalls 11. Von diesem Betrag erklärt der erste Faktor (bei der Anfangslösung mit insgesamt elf Faktoren) 4,879 und damit 44,4% der Gesamtstreuung. Der zweite Faktor erklärt absolut 1,389. Dies entspricht einem

Anteil von 12,6% der Gesamtstreuung. Die Spalte *Kumulierte %* weist die kumulierten Anteile der erklärten Streuung an der Gesamtstreuung aus. Dort ist zum Beispiel zu erkennen, daß die ersten beiden Faktoren bereits 57% der Gesamtstreuung erklären. Ferner ist abzulesen, daß der zusätzliche Erklärungsbeitrag der weiteren Faktoren rapide abnimmt.

Ein zweiter Begriff, der für die Beurteilung der Faktorextraktion von großer Relevanz ist, ist der Begriff der Kommunalität (Communality). Die Kommunalität gibt an, welchen Betrag der Streuung einer Variablen alle Faktoren zusammen genommen erklären. In Abschnitt 26.2, *Das Modell der Faktorenanalyse*, S. 641 wurde gezeigt, daß man für jede der in das Faktormodell aufgenommenen Variablen eine Gleichung in der Art einer multiplen Regressionsgleichung schreiben kann, in der die Variable die abhängige Größe und die Faktoren die erklärenden Größen darstellen. Wie bei einer Regressionsanalyse können die erklärenden Größen (hier also die Faktoren) einen kleinen oder einen großen Teil der Streuung der abhängigen Variablen erklären. Da die Werte der Variablen in Z-Werte transformiert wurden, hat jede Variable eine Streuung (Standardabweichung) von 1. Daher kann auch die für eine Variable ausgewiesene Kommunalität nicht größer als 1 sein. Der Wert 1 wird genau dann erreicht, wenn die Streuung einer Variablen restlos durch die Faktoren erklärt wird.

Die Tabelle in Abbildung 26.7 gibt in der Spalte *Anfänglich* die Kommunalitäten an, die sich ergeben, solange - im ersten Schritt der Faktorextraktion - insgesamt elf Faktoren unterschieden werden. Für alle elf Variablen wird eine Kommunalität von 1 ausgewiesen. Sofern also elf Faktoren gebildet werden, lassen sich die Variablen vollständig durch diese Faktoren erklären. Dennoch ist diese Lösung absolut unbrauchbar, da die elf Variablen durch ebensoviele Faktoren ersetzt wurden, so daß noch keine Verringerung der Komplexität erzielt werden konnte.

Kommunalitäten

	Anfänglich	Extraktion
CMMT	1,000	,573
FEWTW	1,000	,584
FRKONZ	1,000	,581
MANNZTW	1,000	,606
MOE	1,000	,399
PETGTTW	1,000	,684
PETOFTW	1,000	,720
PETSFGTW	1,000	,536
PETWETW	1,000	,719
PETZFGTW	1,000	,562
PROZSUM	1,000	,302

Extraktionsmethode: Hauptkomponentenanalyse

Abbildung 26.7: Kommunalitäten für Lösungen mit elf („Anfänglich“) und zwei („Extraktion“) Faktoren

Die Spalte *Extraktion* gibt die Kommunalitäten für die abschließende Lösung der Faktorextraktion wieder, die alle deutlich kleiner als 1 sind. Dies erklärt sich da-

durch, daß in der abschließenden Lösung nur noch zwei Faktoren im Faktormodell verwendet wurden. Mit diesen beiden Faktoren kann nur ein Teil der Streuung einer jeden Variablen erklärt werden. Dafür sind jedoch zwei Faktoren wesentlich leichter zu handhaben als elf Variablen.

Die Werte der Kommunalitäten liegen stets zwischen 0 und 1. Werte in der Nähe von 0 zeigen an, daß die Faktoren in ihrer Gesamtheit fast nichts von der Streuung der betreffenden Variablen erklären. Werte in der Nähe von 1 bedeuten dagegen, daß durch die Faktoren fast die gesamte Streuung der Variablen erklärt wird. Ist dies der Fall, so haben die Einzelrestfaktoren eine geringe Bedeutung. Liegt die Kommunalität dagegen nahe bei 0, kommt dem Einzelrestfaktor eine große Bedeutung zu.

Kommunalität und *Eigenwert* können leicht verwechselt werden, da sie beide den Betrag einer erklärten Streuung angeben. Dennoch besteht ein deutlicher Unterschied zwischen beiden Größen: Während der Eigenwert angibt, welcher Teil der Gesamtstreuung **aller Variablen** durch **einen bestimmten Faktor** erklärt wird, gibt die Kommunalität an, welcher Teil der Streuung **einer Variablen** durch **alle Faktoren**, die im jeweiligen Modell berücksichtigt wurden, erklärt wird.

Bestimmung der Faktorenzahl

Es wurde bereits deutlich, daß es sinnlos wäre, so viele Faktoren im Modell zu berücksichtigen, wie es Variablen gibt. Andererseits geht aus der Tabelle in Abbildung 26.7 bereits hervor, daß der Anteil der durch die Faktoren erklärten Streuungen der einzelnen Variablen mit abnehmender Faktorenzahl ebenfalls sinkt. Es stellt sich also die Frage, wie viele Faktoren in dem Modell berücksichtigt werden sollen. Diese Frage kann nicht allein anhand einer starren Formel entschieden werden. Vielmehr gilt es, die Anzahl der Faktoren auszuwählen, bei der noch ein hinreichend großer Teil der Streuungen erklärt wird, und die zugleich eine ausreichend große Reduzierung der Komplexität erzielt. Hierbei können folgende Überlegungen helfen: Wenn der Eigenwert eines Faktors kleiner als 1 ist, erklärt dieser Faktor einen geringeren Betrag der Gesamtstreuung als jede einzelne Variable, denn jede Variable erklärt ja immerhin sich selbst und damit eine Streuung von 1 (beachten Sie die erfolgte Transformation in Z-Werte). Aus diesem Grund beläßt SPSS in der Voreinstellung nur solche Faktoren im Modell, die einen Eigenwert von mehr als 1 aufweisen.²⁹³ Im vorliegenden Fall sind dies zwei Faktoren, die in Abbildung 26.6, S. 648 in der rechten Tabellenhälfte ausgewiesen werden.

Diese von SPSS durchgeführte Auswahl der Faktoren stellt jedoch nur eine Voreinstellung dar, die Sie manuell verändern können. Als Entscheidungshilfe hierzu

²⁹³ Diese Voreinstellung können Sie in dem Dialogfeld der Schaltfläche *Extraktion* ändern, siehe im einzelnen Abschnitt 26.8.3, *Extraktionsmethode*, S. 665.

wird häufig eine als *Screepplot*²⁹⁴ bezeichnete Grafik herangezogen. Abbildung 26.8 zeigt das Screepplot für das vorliegende Beispiel. Die Grafik führt die nach der Größe ihrer Eigenwerte geordneten Faktoren auf und zeigt für jeden Faktor die Höhe des Eigenwertes an. Das Screepplot soll (um im Bild zu bleiben) dazu dienen, das Geröll, das sich am Fuße des Berghangs ansammelt, von den wirklich wirksamen Faktoren zu trennen. Typischerweise weist die Kurve im Screepplot das in Abbildung 26.8 zu erkennende Bild auf: Zunächst fällt die Kurve sehr steil ab, weist dann jedoch sehr bald einen Knick auf, um im weiteren Verlauf nur noch sehr schwach abzufallen. Als Faustregel gilt die Empfehlung, die Anzahl an Faktoren zu wählen, bei der die Kurve den Knick aufweist. In diesem Beispiel würde man sich also auch anhand des Screepplot für eine Lösung mit zwei, möglicherweise auch mit drei Faktoren entscheiden.

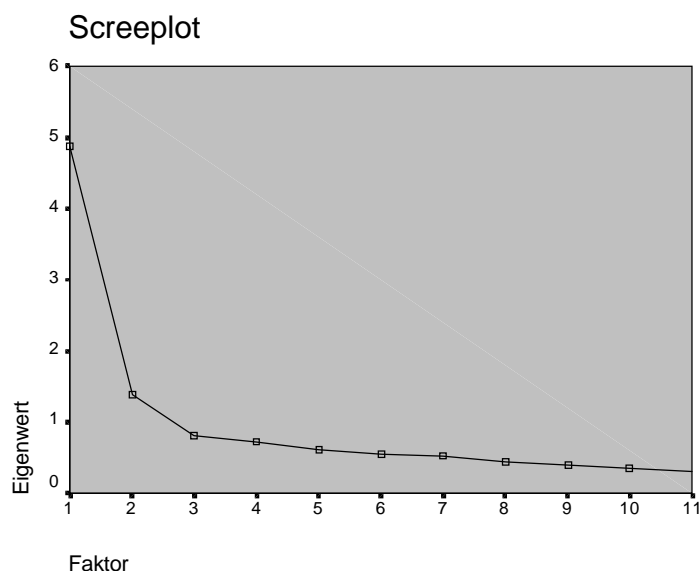


Abbildung 26.8: Screepplot zur Auswahl der optimalen Faktorenanzahl

Faktorladungen

Die gefundenen Faktoren werden für den Analytiker erst dann wertvoll, wenn ihre Relationen zu den einzelnen zu erklärenden Variablen aufgezeigt werden können. Dies geschieht in der Faktormatrix, die für das vorliegende Beispiel in Abbildung 26.9 wiedergegeben wird.

²⁹⁴ Der Begriff wird hier nicht übersetzt, weil auch im Deutschen der englische Ausdruck üblich ist. *Scree* bedeutet soviel wie „Geröll“ oder „Geröllhang“ und soll die typische Form der Grafik, wie sie auch in Abbildung 26.8 dargestellt ist, zum Ausdruck bringen.

Komponentenmatrix^a

	Komponente	
	1	2
CMMT	,754	,073
FEWTW	,719	-,258
FRKONZ	,676	-,352
MANNZTW	,685	-,370
MOE	,615	,143
PETGTTW	,709	,426
PETOFTW	,795	-,297
PETSGFTW	,729	-,066
PETWETW	,608	,591
PETZFGTW	,447	,602
PROZSUM	,503	-,222

Extraktionsmethode: Hauptkomponentenanalyse

a. 2 Komponenten extrahiert

Abbildung 26.9: Faktormatrix (Komponentenmatrix) mit dem Zusammenhang zwischen den Faktoren und den einzelnen Variablen

Die Faktormatrix gibt für jede Variable des Faktormodells die Koeffizienten an, mit denen die beiden Faktoren in die Gleichung zur Erklärung der Variablen eingehen. Diese Koeffizienten werden meistens als *Faktorladungen* bezeichnet; entsprechend ist für die Matrix auch die Bezeichnung *Faktorladungsmatrix* üblich. Aus der Matrix läßt sich zum Beispiel für die Variable *fewtw* ablesen, daß diese durch folgende Gleichung beschrieben werden kann:

$$\text{fewtw} = 0,719 \cdot F_1 - 0,258 \cdot F_2$$

An der absoluten Größe einer Faktorladung können Sie die Bedeutung des jeweiligen Faktors für die betreffende Variable ablesen. So hat Faktor 2 eine sehr große Bedeutung für die Variable *petzfgtw*, aber nur eine sehr geringe für die Variable *cmmt*. Demgegenüber wird *cmmt* stark durch den ersten Faktor erklärt.

Reproduzierte Korrelationskoeffizienten

Bisher wurde implizit unterstellt, die Faktoren seien nicht miteinander korreliert. Wenn diese Annahme tatsächlich zutrifft, spricht man von *Orthogonalität* der Faktoren. Die Koeffizienten (Ladungen) der Faktoren wären dann größengleich mit den Korrelationskoeffizienten zwischen den Faktoren und der Variablen. Im nächsten faktoranalytischen Schritt, der *Rotation*, wird sich jedoch herausstellen, daß die Annahme orthogonaler Faktoren aufgegeben werden kann.

Soweit die grundlegende faktoranalytische Ausgangsannahme, daß die zwischen den Variablen beobachteten Korrelationen auf einen gemeinsamen Hintergrundfaktor zurückgehen, zutrifft, müssen sich die Korrelationskoeffizienten zwischen Variablenpaaren aus den Faktorladungen schätzen lassen. Allgemein gilt die folgende Beziehung:

$$r_{ij} = \sum_{f=1}^k r_{fi} \cdot r_{fj}$$

Dabei gibt k die Anzahl der Faktoren an, und r_{fi} bzw. r_{fj} bezeichnen den Korrelationskoeffizienten des f -ten Faktors mit der i -ten bzw. der j -ten Variablen. Für den Korrelationskoeffizienten zwischen den Variablen *petgttw* und *cmmt* berechnet sich die Schätzung zum Beispiel folgendermaßen:²⁹⁵

$$r_{\text{petgttw, cmmt}} = 0,754 \cdot 0,709 + 0,073 \cdot 0,426 = 0,566$$

Der Wert 0,566 stellt eine Schätzung für den Korrelationskoeffizienten zwischen den beobachteten Variablen *petgttw* und *cmmt* dar. Dieser Korrelationskoeffizient hätte natürlich auch direkt aus den Ursprungswerten errechnet werden können, was ja auch tatsächlich bereits geschah: Die Ergebnisse für alle möglichen Korrelationskoeffizienten werden oben in der Korrelationsmatrix in Abbildung 26.2, S. 644 aufgeführt. Dort steht für die Variablenkombination *petgttw/cmmt* der Wert 0,524. Die Abweichung zwischen dem Originalwert und dem geschätzten Wert beträgt damit -0,052. Diese Abweichung wird in Analogie zum Regressionsmodell *Residuum* genannt. Der mit Hilfe der Faktoren geschätzte Wert wird auch als *reproduzierter Korrelationskoeffizient* bezeichnet.

Reproduzierte Korrelationen

	CMMT	FEWTW	FRKONZ	MANNZTW	MOE	PETGTTW	PETOFTW	PETSFGTW	PETWETW	PETZFGTW	PROZSUM	
Reproduzierte Korrelation	CMMT	,573 ^b	,523	,484	,489	,474	,565	,577	,545	,502	,381	,363
	FEWTW	,523	,584 ^b	,577	,589	,406	,400	,649	,542	,285	,166	,419
	FRKONZ	,484	,577	,581 ^b	,594	,365	,329	,642	,516	,203	,090	,418
	MANNZTW	,489	,589	,594	,606 ^b	,369	,328	,655	,524	,198	,084	,427
	MOE	,474	,406	,365	,369	,399 ^b	,497	,446	,439	,459	,361	,277
	PETGTTW	,565	,400	,329	,328	,497	,684 ^b	,437	,489	,683	,573	,262
	PETOFTW	,577	,649	,642	,655	,446	,437	,720 ^b	,600	,308	,176	,466
	PETSFGTW	,545	,542	,516	,524	,439	,489	,600	,536 ^b	,404	,286	,381
	PETWETW	,502	,285	,203	,198	,459	,683	,308	,404	,719 ^b	,628	,175
	PETZFGTW	,381	,166	,090	,084	,361	,573	,176	,286	,628	,562 ^b	,091
	PROZSUM	,363	,419	,418	,427	,277	,262	,466	,381	,175	,091	,302 ^b
	Residuum ^a	CMMT		-,024	-,056	,006	-,061	-,042	-,062	-,038	-,076	-,051
FEWTW		-,024		-,053	-,083	,021	-,067	-,121	-,009	,008	-,123	
FRKONZ		-,056	-,053		-,107	-,041	,007	-,061	-,036	,033	,046	-,149
MANNZTW		,006	-,083	-,107		-,068	,011	-,068	-,077	-,007	,091	-,068
MOE		-,061	,021	-,041	-,068		-,127	-,028	-,079	-,039	-,104	-,048
PETGTTW		-,042	,006	,007	,011	-,127		-,037	-,016	-,042	-,151	,006
PETOFTW		-,062	-,067	-,061	-,068	-,028	-,037		-,027	,012	,052	-,015
PETSFGTW		-,038	-,121	-,036	-,077	-,079	-,016	-,027		-,042	-,026	-,053
PETWETW		-,076	-,009	,033	-,007	-,039	-,042	,012	-,042		-,189	,071
PETZFGTW		-,051	,008	,046	,091	-,104	-,151	,052	-,026	-,189		,014
PROZSUM		-,116	-,123	-,149	-,068	-,048	,006	-,015	-,053	,071	,014	

Extraktionsmethode: Hauptkomponentenanalyse

a. Residuen werden zwischen beobachteten und reproduzierten Korrelationen berechnet. Es gibt 27 (49,0%) nichtredundante Residuen mit Absolutwerten > 0,05.

b. Reproduzierte Kommunalitäten

Abbildung 26.10: Matrix der reproduzierten Korrelationskoeffizienten

In der oberen Hälfte der Tabelle aus Abbildung 26.10 ist die Matrix der reproduzierten Korrelationskoeffizienten für alle Variablenpaare wiedergegeben. Dies ist

²⁹⁵ Die Korrelationskoeffizienten zwischen den Faktoren und den Variablen können Sie der Faktormatrix aus Abbildung 26.9 entnehmen.

eine symmetrische Matrix, in deren Hauptdiagonale die Kommunalitäten der Variablen aufgeführt werden.

Unterhalb der Tabelle wird mitgeteilt, daß 27 Residuen (dies sind 49% aller Residuen) einen Wert über 0,05 haben. Diese Angabe kann als ein grobes Maß für den *Fit* (die Anpassung) des Modells angesehen werden. Darüber hinaus geben natürlich auch die einzelnen Residuen selbst (untere Tabellenhälfte) darüber Aufschluß, wie gut die einzelnen Variablenkorrelationen durch die extrahierten Faktoren erfaßt werden.

Verschiedene Methoden der Faktorextraktion

Für die Faktorextraktion stehen außer der im Beispiel verwendeten Methode der Hauptkomponentenanalyse noch weitere Verfahren zur Verfügung. Diese Verfahren unterscheiden sich in der Art, in der die beste Modellanpassung erreicht wird:

- **Hauptkomponenten:** Dieses Verfahren ist voreingestellt und wurde im bisherigen Beispiel verwendet.
- **Hauptachsen-Faktorenanalyse:** Dieses Verfahren ist der Hauptkomponentenmethode recht ähnlich. Der Unterschied zu dieser besteht darin, daß bei der Hauptachsenmethode in die Diagonale der Korrelationsmatrix als Schätzungen für die Kommunalitäten in einem ersten Schritt die quadrierten multiplen Korrelationskoeffizienten eingesetzt werden. Auf dieser Basis werden dann die geeignete Faktorenanzahl und die daraus resultierenden Kommunalitäten berechnet. Diese werden in einem zweiten Schritt zum Ausgangspunkt einer erneuten Faktorenschätzung gemacht, was wiederum zu neuen Kommunalitäten führt usw. Der Iterationsprozeß wird so lange fortgesetzt, bis sich die Kommunalitäten nicht mehr nennenswert ändern.
- **Ungewichtete kleinste Quadrate:** Hierbei wird in einem Iterationsprozeß für eine feste Faktorenanzahl eine Faktorladungsmatrix berechnet, für die die Summe der quadrierten Differenzen aus Korrelationskoeffizienten und reproduzierten Korrelationskoeffizienten (also die Summe der quadrierten Residuen) minimal ist.
- **Verallgemeinerte kleinste Quadrate:** Diese Form der Kleinstquadratmethode verfährt im Grunde genauso wie die vorhergehende, gewichtet aber die Korrelationskoeffizienten invers mit den Einzelrestfaktoren.
- **Maximum Likelihood:** Mit dieser Methode werden solche Parameter geschätzt, für die die Wahrscheinlichkeit, die beobachtete (Stichproben-)Korrelationsmatrix hervorzubringen, am größten ist. Dabei wird unterstellt, daß die Stichprobe einer multivariaten Normalverteilung entstammt. Auch bei dieser Methode werden die Korrelationen so wie bei der verallgemeinerten kleinsten Quadrate-Methode gewichtet.
- **Alpha-Faktorisierung:** Darüber hinaus stehen noch die Alpha- und die
- **Image-Faktorisierung** zur Verfügung, auf die hier aber nicht näher eingegangen wird.

26.6 Rotation

Zweck der Rotation

Im letzten Schritt wurden zwei Faktoren identifiziert, die nach dem Kriterium ihres Eigenwertes eine gewisse Erklärungskraft besitzen. Bisher wurden diese Faktoren noch nicht interpretiert, sondern als rein rechentechnische Ergebnisse betrachtet. Die Faktorenanalyse bleibt jedoch ohne Aussagekraft, wenn die Bedeutung der Faktoren nicht inhaltlich bestimmt werden kann. Die Faktoren müssen also interpretiert werden.

Die Interpretation der Faktoren ergibt sich aus der Relation, die sie zu den Beobachtungsvariablen haben, für die sie ja die Hintergrundvariablen darstellen sollen. Die Beziehungen der Faktoren zu den einzelnen Variablen können der Faktorladungsmatrix entnommen werden: Große Faktorladungen zeigen eine große, kleine dagegen eine geringe Bedeutung eines Faktors für die entsprechende Variable an. Ein Faktor ist dann leicht zu interpretieren, wenn einige Variablen, die untereinander eine homogene Bedeutung haben, hoch auf ihn laden und gleichzeitig die Ladungen der anderen Variablen auf diesen Faktor gering sind. Umgekehrt fällt eine sinnvolle Interpretation jedoch sehr schwer oder ist vollkommen unmöglich, wenn ein Faktor zu sämtlichen Variablen des Modells eine relativ starke Korrelation aufweist. So zeigt zum Beispiel die Faktorladungsmatrix aus Abbildung 26.9, S. 652, daß der erste Faktor für zehn der insgesamt elf Variablen eine Faktorladung von mindestens 0,5 aufweist, die bei vielen Variablen sogar deutlich darüber liegt. Damit läßt sich dieser Faktor kaum interpretieren, weil er (anscheinend) sehr viele heterogene Variablen erklärt. Eine solche Situation ist für den ersten Versuch einer Faktorladungsmatrix nicht untypisch. Zur leichteren Interpretation sind daher verschiedene Verfahren entwickelt worden, die die Faktorladungsmatrix einer als Rotation bezeichneten Transformation unterwerfen. Der Begriff Rotation erklärt sich daraus, daß bei der Transformation die Achsen des Koordinatensystems, in dem die Faktorladungen dargestellt sind, gedreht werden. Dies sei im folgenden an den beiden Faktoren, deren Ladungen sich in einem zweidimensionalen Diagramm darstellen lassen, demonstriert.

In dem Ladungsplot aus Abbildung 26.11, das mit dem gleichnamigen Kontrollkästchen im Dialogfeld *Rotation* der Faktorenanalyse erzeugt wurde, werden die Koeffizienten der Faktormatrix (die Ladungen) für die beiden Faktoren grafisch dargestellt. Die Koordinaten der Punkte entsprechen den Faktorladungen der einzelnen Variablen auf die beiden Faktoren, die oben bereits in der Faktorladungsmatrix aus Abbildung 26.9 angegeben wurden. Die beiden Faktoren wären relativ einfach zu interpretieren, wenn sich die Punkte in dem Diagramm eng um die Achsen scharten und gleichzeitig weit vom Ursprung entfernt lägen, denn dies würde bedeuten, daß jede Variable für den einen Faktor eine große, und für den anderen eine geringe Ladung hätte. In dem Diagramm aus Abbildung 26.11 trifft dies für acht der elf Punkte zu, denn diese streuen relativ eng um die waagerechte Achse, die den Faktor 1 skaliert. Diese acht Punkte repräsentieren Variablen, die relativ stark auf Faktor 1 und nur gering auf Faktor 2 laden.

Die drei Punkte für die Variablen *petwetw*, *petgttw* und *petzfgtw* liegen dagegen eher in der Mitte des ersten Quadranten. Wenn allerdings das Achsenkreuz um ca. 30° gedreht würde, ergäbe sich ungefähr das angestrebte Bild, bei dem sich die Punkte um die Achsen scharen. Grafisch kann man sich eine solche Achsendrehung kaum für mehr als zwei Achsen vorstellen, rechnerisch ist sie jedoch für ein n-dimensionales Koordinatensystem durchführbar. Eine solche *Rotation* der Achsen ändert natürlich die Werte der Faktorkoeffizienten; dies ist ja schließlich geradezu das Ziel der Rotation. Sie verändert aber in keiner Weise den Fit des Modells. Das bedeutet vor allem, daß weder die Kommunalitäten noch die Eigenwerte durch die Rotation geändert werden.

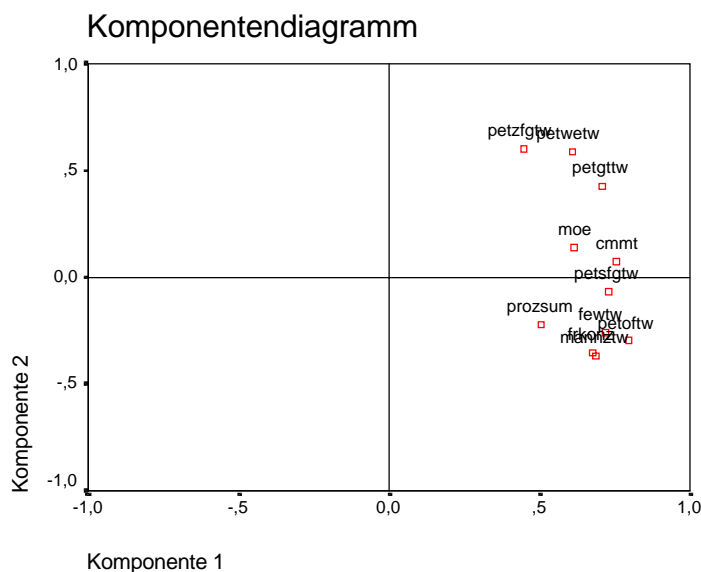


Abbildung 26.11: Ladungsplot für die beiden Faktoren „Faktor 1“ und „Faktor 2“

Rotationsmethoden

Bei SPSS stehen verschiedene Rotationsmethoden zur Verfügung. Dabei ist zunächst der Unterschied zwischen orthogonaler (rechtwinkliger) und schiefwinkliger Rotation von Bedeutung: Bei den Verfahren mit orthogonaler Rotation werden die Achsen lediglich gedreht, die relative Position der Achsen zueinander bleibt jedoch unverändert, so daß die Achsen im rechten Winkel zueinander bleiben. Dem entspricht inhaltlich die Modellvoraussetzung, daß die Faktoren nicht miteinander korreliert sind, sich also orthogonal zueinander verhalten. Bei einer schiefwinkligen (oblique) Rotation wird diese Annahme dagegen aufgegeben. Es wird also eine Korrelation der Faktoren zugelassen, so daß die Koordinatenachsen nicht mehr im rechten Winkel zueinander stehen.

SPSS stellt drei Verfahren der orthogonalen Rotation zur Verfügung. Bei der *Varimax*-Methode werden die Achsen so rotiert, daß die Anzahl von Variablen mit hoher Faktorladung minimiert wird. Dies ist wohl das gebräuchlichste Verfahren, durch das vor allem die Interpretierbarkeit der Faktoren erhöht wird. Mit der *Quartimax*-Methode wird die Zahl der Faktoren minimiert, die zur Erklärung einer Variablen erforderlich sind. Damit ist dieses Verfahren in erster Linie auf eine leichtere Interpretierbarkeit der Variablen ausgerichtet. Allerdings resultiert dieses Verfahren häufig nur in einem allgemeinen und umfassenden Faktor, der zu den meisten Variablen eine nennenswerte Ladung aufweist. Schließlich wird mit der *Equamax*-Methode ein Rotationsverfahren bereitgestellt, das eine Mischung aus der *Varimax*- und der *Quartimax*-Methode darstellt.

Für die schiefwinklige Rotation stehen zwei Verfahren zur Verfügung, das *Direkte Oblimin* und die *Promax*-Rotation. Beim Direkten Oblimin kann der Grad der Schiefwinkligkeit wahlweise explizit vorgegeben oder von SPSS bestimmt werden. Die Promax-Rotation bietet sich vor allem für umfangreiche Stichproben an, da die Rotation hierbei einen deutlich geringeren Rechenaufwand erfordert als bei dem Direkten Oblimin.

Abbildung 26.12 zeigt die Faktorladungsmatrizen für eine Varimax-Rotation (linke Tabelle) und für eine schiefwinklige Rotation (rechte Tabelle). Die entsprechenden Ladungsplots sind in den Abbildungen 26.13 (Varimax-Rotation) und 26.14 (schiefwinklige Rotation) dargestellt. Um die Matrix und den Ladungsplot für eine Varimax-Rotation zu erstellen, wählen Sie in dem Dialogfeld *Rotation* die Optionen *Varimax*, *Rotierte Lösung* und *Ladungsdiagramm(e)*. Für die schiefwinklige Rotation wählen Sie entsprechend die Optionen *Oblimin direkt*, *Rotierte Lösung* und *Ladungsdiagramm(e)*.

Rotierte Komponentenmatrix ^a			Mustermatrix ^a		
	Komponente			Komponente	
	1	2		1	2
CMMT	,571	,497	CMMT	,516	,379
FEWTW	,736	,208	FEWTW	,753	,026
FRKONZ	,755	,106	FRKONZ	,795	-,088
MANNZTW	,773	,097	MANNZTW	,816	-,102
PETGTTW	,330	,758	PETGTTW	,202	,722
PETOFTW	,820	,220	PETOFTW	,842	,017
PETSFGTW	,632	,369	PETSFGTW	,608	,227
PETWETW	,152	,834	PETWETW	-,007	,851
PETZFGTW	,014	,749	PETZFGTW	-,138	,797
PROZSUM	,538	,111	PROZSUM	,559	-,024
MOE	,418	,474	MOE	,355	,395

Extraktionsmethode: Hauptkomponentenanalyse.
 Rotationsmethode: Varimax mit Kaiser-Normalisierung.
 a. Die Rotation ist in 3 Iterationen konvergiert.

Extraktionsmethode: Hauptkomponentenanalyse.
 Rotationsmethode: Oblimin mit Kaiser-Normalisierung.
 a. Die Rotation ist in 6 Iterationen konvergiert.

Abbildung 26.12: Faktorladungsmatrizen nach einer Varimax-Rotation (linke Tabelle) und nach einer schiefwinkligen Rotation (Direktes Oblimin; rechte Tabelle)

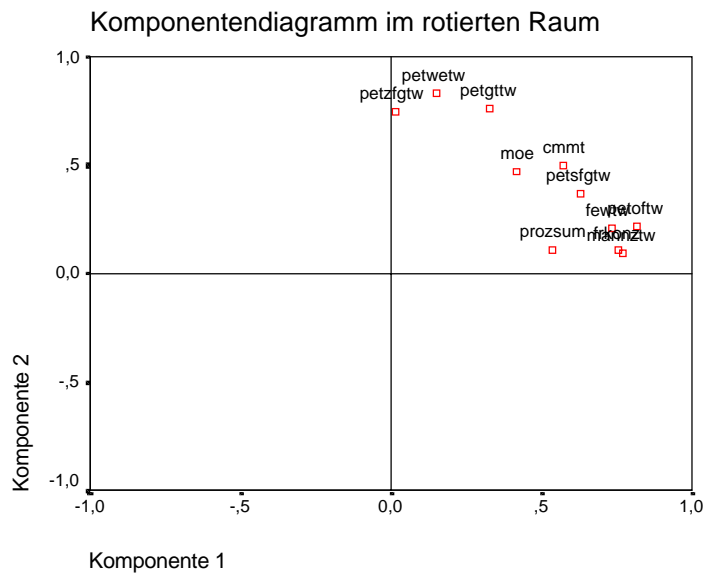


Abbildung 26.13: Ladungsplot nach der Varimax-Rotation

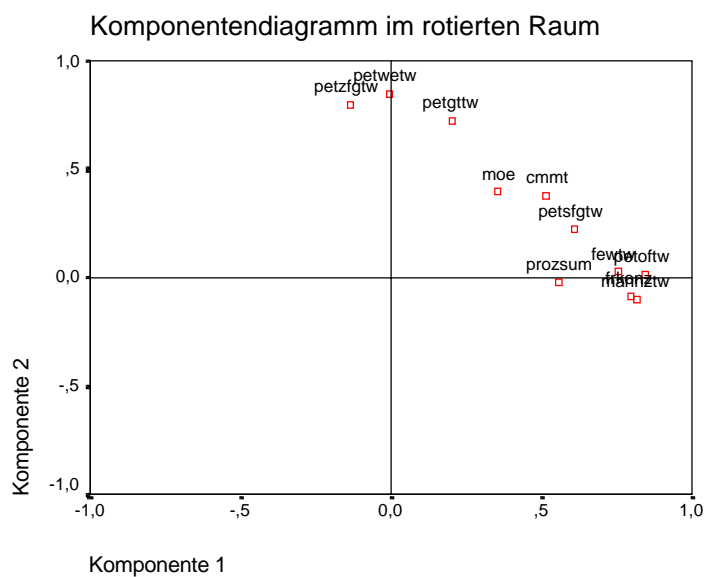


Abbildung 26.14: Ladungsplot nach der Direkten Oblimin-Rotation

Ein Vergleich der Matrizen und Ladungsplots für die rotierten Lösungen mit den nicht rotierten (Abbildung 26.9, S. 652 und Abbildung 26.11, S. 656) zeigt, daß

die Faktoren in den rotierten Formen klarerer Beziehungen zu den einzelnen Variablen aufweisen. Insbesondere bei der schiefwinkligen Rotation sind für die einzelnen Faktoren deutlich ausgeprägte Unterschiede zwischen sehr hohen und sehr geringen Faktorladungen zu erkennen.

Sortierte Darstellung der Faktorladungen

Die Ladungsplots erleichtern oftmals die Interpretation der Faktoren, allerdings sind sie auf die gleichzeitige Darstellung zweier Faktoren beschränkt. Dies stellt zwar im vorliegenden Beispiel keine Einschränkung dar, das wäre aber anders, wenn mehr als zwei Faktoren identifiziert worden wären. Um den Blick auf das Wesentliche trotz des zum Teil sehr umfangreichen Output der Faktorenanalyse zu erleichtern, bietet SPSS die Möglichkeit, die Faktorladungen in einer nach der Größe sortierten Form ausgeben zu lassen, so daß Variablen, die auf denselben Faktor hohe Ladungen aufweisen, gemeinsam als Gruppe erscheinen. Eine solche Darstellung läßt sich auch noch mit einer zusätzlichen Anweisung kombinieren, die bewirkt, daß Ladungen nicht ausgewiesen werden, wenn sie einen vom Benutzer definierten Grenzwert unterschreiten. Im folgenden Beispiel wurde der Grenzwert 0,4 gewählt. Diese Angaben nehmen Sie in dem Dialogfeld *Optionen* vor. Dem Beispiel liegt ferner - im Unterschied zu den bisherigen Ergebnissen, die auf der Hauptkomponentenmethode basierten - die Extraktionsmethode *Hauptachsen* zugrunde (diese wählen Sie in dem Dialogfeld *Extraktion*). Die Koeffizienten der Varimax-Rotation sind in Abbildung 26.15 wiedergegeben.

Rotierte Komponentenmatrix^a

	Komponente	
	1	2
PETOFTW	,808	
MANNZTW	,705	
FRKONZ	,679	
FEWTW	,676	
PETSFGTW	,588	
CMMT	,552	,456
PROZSUM	,428	
MOE	,401	
PETWETW		,810
PETGTTW		,719
PETZFGTW		,529

Extraktionsmethode: Hauptkomponentenanalyse.
Rotationsmethode: Varimax mit Kaiser-Normalisierung.

a. Die Rotation ist in 3 Iterationen konvergiert.

Abbildung 26.15: Faktorladungsmatrix, erstellt mit der Extraktionsmethode „Hauptachsen-Faktorenanalyse“, nach einer Varimax-Rotation; sortierte Koeffizienten $\geq 0,4$

Diese Aufbereitung der Faktorkoeffizienten erleichtert zumindest die formale Interpretation der Faktoren erheblich, weil sich ein äußerlich klares Bild ergibt: Die drei Variablen *petwetw*, *petgttw* und *petzfgtw* besitzen nur für den Faktor 2 Faktorladungen $\geq 0,4$, die anderen Variablen nur für den Faktor 1. Die einzige Aus-

nahme bildet *cmmt*, die für beide Faktoren Ladungen $> 0,4$ aufweist. Die drei Variablen *petwetw*, *petgtw* und *petzftw* geben inhaltlich Testergebnisse zur sprachlichen Intelligenz wieder (vgl. die Tabelle in Abbildung 26.1, S. 640), so daß Faktor 2 als *Sprachliche Intelligenz (SI)* interpretiert werden kann. Die anderen Variablen geben Testergebnisse zur allgemeinen Intelligenz wieder. Faktor 1 läßt sich somit als *Allgemeine Intelligenz (AI)* interpretieren.

Ein Vergleich der Faktorkoeffizienten aus der Varimax-Rotation auf Basis der Hauptachsenmethode (Abbildung 26.15) mit den entsprechenden Koeffizienten, die mit der Hauptkomponentenmethode gefunden wurden (Abbildung 26.9, S. 652 und Abbildung 26.12, S. 657), zeigt, daß sowohl die Extraktions- als auch die Rotationsmethode einen erheblichen Einfluß auf die Faktorladungen haben, der sich nicht zuletzt auf die Interpretation der Faktoren auswirkt.

Besonderheiten der schiefwinkligen Rotation

In den vorangehenden Beispielen wurde die schiefwinklige Rotation quasi gleichberechtigt neben der orthogonalen Rotation nach der Varimax-Methode vorgestellt. Dieses Vorgehen sollte allerdings nicht darüber hinwegtäuschen, daß Sie durch die schiefwinklige Rotation die zentrale Eigenschaft des ursprünglichen Faktorenmodells aufgeben, nach der die Faktoren miteinander keinerlei Korrelation aufweisen. Diese Eigenschaft garantiert die klare Trennung der Faktoren voneinander. Bei der schiefwinkligen Rotation überlappen sich die Faktoren dagegen in dem Sinne, daß sie miteinander korrelieren können. Dies ist der Preis dafür, daß man auf Basis einer schiefwinkligen Rotation oft zu einer eindeutigeren Interpretation der Faktoren gelangt als bei orthogonaler Rotation. Eine gewisse Rechtfertigung für das Aufgeben der Nicht-Korrelationseigenschaft der Faktoren mag allerdings darin gesehen werden, daß komplexe Hintergrundvariablen auch in der konkreten Realwelt (Nicht-Modellwelt) im allgemeinen miteinander korrelieren werden; die zur Verfügung stehenden theoretischen und praktischen Begriffe zur Beschreibung komplexer Phänomene dürften nur selten so trennscharf sein, daß es keinen Bereich der Überlappung gibt.

In jedem Fall ist noch auf die folgende formale Besonderheit der schiefwinkligen Rotation hinzuweisen: Obwohl die Kommunalitäten durch die schiefwinklige Rotation unverändert bleiben - wie dies auch bei der orthogonalen Rotation der Fall ist - gilt eine andere Eigenschaft nicht mehr, die bei der orthogonalen Rotation erhalten bleibt: Dort sind Matrizen der Faktorladungen und der Korrelationskoeffizienten zwischen Faktoren und Variablen identisch, nach der schiefwinkligen Rotation gilt dies nicht mehr. Aus diesem Grunde gibt SPSS bei der schiefwinkligen Rotation außer der Faktorladungsmatrix (die dann *Mustermatrix* heißt) auch die Matrix der Korrelationskoeffizienten (*Strukturmatrix*) aus, die oben nicht dargestellt wurde. Darüber hinaus wird bei der schiefwinkligen Rotation, weil dies nun eine relevante Information ist, eine Tabelle mit den Korrelationskoeffizienten der Faktoren ausgegeben.

26.7 Schätzung von Faktorwerten

Schätzmethoden

Für viele faktoranalytische Untersuchungen ist das Ziel erreicht, wenn eine befriedigende Interpretation der Faktoren gelungen ist. Häufig ist man aber auch daran interessiert, die Faktoren als Variablen in weiteren Untersuchungen verwenden zu können. Für solche Zwecke ist es erforderlich, über konkrete Faktorwerte zu verfügen. Außer bei der Verwendung der Hauptkomponentenmethode als Extraktionsverfahren, bei der Faktorwerte exakt berechnet werden können, müssen die Faktorwerte geschätzt werden. Hierzu kann der Anwender bei SPSS zwischen drei verschiedenen Schätzverfahren wählen. Alle drei Methoden (*Regression*, *Bartlett* und *Anderson-Rubin*) führen zu Faktorwerten mit einem Mittelwert von null. Die Anderson-Rubin-Methode ergibt unkorrelierte Werte mit einer Standardabweichung von 1, und zwar auch dann, wenn für die Faktoren eine Korrelation geschätzt wurde. Faktorwerte, die nach der Regressionsmethode geschätzt wurden, weisen eine Varianz auf, die sich aus der quadrierten multiplen Korrelation zwischen den geschätzten und den wahren Faktorwerten ergibt. Die Faktorwerte nach dieser Methode können auch dann korreliert sein, wenn die Faktoren orthogonal sind. Im übrigen führen alle drei Schätzverfahren bei einer Faktorextraktion nach der Hauptkomponentenmethode zu denselben (nicht geschätzten, sondern berechneten) Ergebnissen.²⁹⁶ Die Schätzmethoden können Sie in dem Dialogfeld der Schaltfläche *Werte* auswählen.

Speicherung der Faktorwerte

Wenn Sie für weitere analytische Zwecke über Faktorwerte verfügen möchten, können Sie für jeden Faktor eine neue Variable mit den Faktorwerten in die Datendatei einfügen lassen. Kreuzen Sie hierzu die Option *Als Variablen speichern* in dem Dialogfeld der Schaltfläche *Werte* an. SPSS fügt dann zusätzlich einen Hinweis in die Ausgabedatei ein, der die Namen und Label der neu erstellten Variablen angibt. Auch in die Datei *Frühgeborene.sav* wurden Variablen für die beiden im Beispiel gefundenen Faktoren eingefügt. Den Variablen wurden von SPSS die Namen *fac1_1* und *fac2_1* gegeben, diese Namen wurden jedoch geändert, so daß die Variablen nun *ai_werte* und *si_werte* heißen.

Verwendung der Faktorwerte in weiteren Analysen

In der Forschungsarbeit, aus der das Beispiel dieses Kapitels stammt, wurden die Intelligenztests im sechsten Lebensjahr der Frühgeborenen vor allem zu dem Zweck durchgeführt, mögliche Zusammenhänge zwischen perinatalen Faktoren einerseits und der Intelligenzentwicklung andererseits aufzuklären. Daher enthält die Stichprobe auch zahlreiche Variablen mit Informationen, die mit der Geburt

²⁹⁶ Bei der Hauptkomponentenmethode lassen sich die Faktorwerte exakt berechnen, bei allen übrigen Extraktionsverfahren müssen sie dagegen geschätzt werden.

zusammenhängen, unter anderem auch die Variable *o_40*. Diese gibt in klassifizierter Form (Klassen 1, 2 und 3) an, wie lange der frühgeborene Säugling mit Sauerstoff einer bestimmten Konzentration behandelt wurde. Um den Einfluß dieser Variablen auf die mit Hilfe der Faktorenanalyse gewonnenen Faktoren *Allgemeine Intelligenz* und *Sprachliche Intelligenz* zu prüfen, empfiehlt es sich, je eine einfaktorielle Varianzanalyse für die beiden abhängigen Variablen *ai_werte* und *si_werte* mit dem Faktor *o_40* durchzuführen.²⁹⁷ Die Ergebnisse der beiden Varianzanalysen sind in den Abbildungen 26.16 (*Allgemeine Intelligenz*) und 26.17 (*Sprachliche Intelligenz*) wiedergegeben.

ANOVA

AI_WERTE

	Quadratsumme	df	Mittel der Quadrate	F	Signifikanz
Zwischen den Gruppen	11,750	2	5,875	7,343	,001
Innerhalb der Gruppen	239,219	299	,800		
Gesamt	250,969	301			

Descriptives

AI_WERTE

	N	Mittelwert	Standardabweichung	Standardfehler	95% Konfidenzintervall für den Mittelwert		Minimum	Maximum
					Untergrenze	Obergrenze		
1	156	,145	,725	,058	,031	,260	-3,103	1,394
2	113	-,056	,963	,091	-,235	,123	-3,373	1,499
3	33	-,495	1,299	,226	-,956	-,035	-3,880	,980
Gesamt	302	,000	,913	,053	-,103	,103	-3,880	1,499

Abbildung 26.16: Ergebnisse einer einfaktoriellen Varianzanalyse mit „ai_werte“ als abhängiger Variablen und „o_40“ als Faktor

ANOVA

SI_WERTE

	Quadratsumme	df	Mittel der Quadrate	F	Signifikanz
Zwischen den Gruppen	,370	2	,185	,237	,789
Innerhalb der Gruppen	233,399	299	,781		
Gesamt	233,769	301			

Descriptives

SI_WERTE

	N	Mittelwert	Standardabweichung	Standardfehler	95% Konfidenzintervall für den Mittelwert		Minimum	Maximum
					Untergrenze	Obergrenze		
1	156	-,018	,842	,067	-,151	,116	-3,715	1,366
2	113	-,004	,899	,085	-,172	,163	-3,201	1,747
3	33	,098	1,017	,177	-,262	,459	-3,046	1,691
Gesamt	302	,000	,881	,051	-,100	,100	-3,715	1,747

Abbildung 26.17: Ergebnisse einer einfaktoriellen Varianzanalyse mit „si_werte“ als abhängiger Variablen und „o_40“ als Faktor

²⁹⁷ Die Bezeichnung *Faktor* für die Variable *o_40* rührt daher, daß die gruppenbildende Variable in der Varianzanalyse als Faktor bezeichnet wird. Zur einfaktoriellen Varianzanalyse siehe ausführlicher Kapitel 20, *Einfaktorielle ANOVA*.

Es zeigt sich, daß die Variable *o_40* (Sauerstoffbehandlung) die Streuung der Werte des Faktors *Allgemeine Intelligenz* signifikant erklärt. Dagegen ist der entsprechende Zusammenhang für den Faktor *Sprachliche Intelligenz* deutlich insignifikant. Wie weitere varianz- und regressionsanalytische Auswertungen zeigen, läßt sich der Faktor *Sprachliche Intelligenz* dagegen gut durch soziale und kulturelle Einflußfaktoren, die ebenfalls in der Untersuchung erhoben wurden, erklären. Dieses Ergebnis wird von der Forschungsgruppe folgendermaßen gedeutet: Die Sauerstoffbehandlung übt einen schädigenden organischen Einfluß auf das Gehirn der Säuglinge aus, das gerade bei Frühgeborenen besonders wenig entwickelt ist, so daß dadurch die allgemeine Intelligenzentwicklung beeinträchtigt wird. Die sprachliche Intelligenz hängt dagegen besonders von sozialen und kulturellen Faktoren ab, die unabhängig von der Sauerstoffbehandlung sind. Dieses Ergebnis zeigt ein Dilemma in der Behandlung Frühgeborener an: Die Überlebenschance Frühgeborener steigt deutlich durch den Einsatz intensivmedizinischer Maßnahmen, zu denen auch die Sauerstoffbehandlung zählt. Die größere Überlebenschance scheint aber ihren Preis in Form beeinträchtigter Entwicklung der allgemeinen Intelligenz zu haben.

Die Datei *Frühgeborene.sav* enthält zahlreiche weitere perinatale, soziale und kulturelle Variablen, Daher können Sie selbst prüfen, in welcher Weise sich die beiden Faktoren *AI* und *SI* von weiteren Variablen erklären lassen.

26.8 Einstellungen der Faktorenanalyse

26.8.1 Variablen angeben

Um eine Faktorenanalyse durchzuführen, wählen Sie den Befehl



STATISTIK
DIMENSIONSREDUKTION ▶
FAKTORENANALYSE...

Dieser Befehl öffnet das in Abbildung 26.18 wiedergegebene Dialogfeld. Die Faktorenanalyse kann durch eine Reihe von Angaben differenziert gesteuert werden. Dabei ist lediglich die Angabe einer Liste von Variablen, für die die Faktorenanalyse durchgeführt werden soll, obligatorisch. Verschieben Sie diese Variablen in das Feld *Variablen*.

Zusätzlich können Sie eine *Auswahlvariable* angeben, um nur solche Fälle in die Faktorenanalyse einzubeziehen, die in dieser Variablen einen bestimmten ganzzahligen Wert aufweisen. Verschieben Sie die betreffende Variable in das Feld *Auswahlvariable*, und klicken Sie anschließend auf die daneben befindliche Schaltfläche *Wert*. Diese öffnet ein Dialogfeld, daß lediglich ein einziges Eingabefeld enthält. Geben Sie in dieses Feld den ganzzahligen Wert ein, den die in die Faktorenanalyse einzubeziehenden Fälle in der Auswahlvariablen aufweisen.

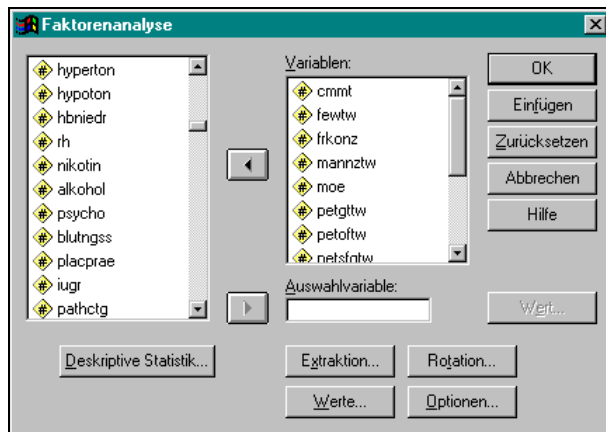


Abbildung 26.18: Dialogfeld des Befehls STATISTIK, DIMENSIONSREDUKTION, FAKTORENANALYSE

26.8.2 Deskriptive Statistiken

Mit der Schaltfläche *Deskriptive Statistik* öffnen Sie das in Abbildung 26.19 dargestellte Dialogfeld, in dem Sie für verschiedene statistische Tabellen und Korrelationsmatrizen wählen können, ob diese im Output der Faktorenanalyse ausgegeben werden sollen.

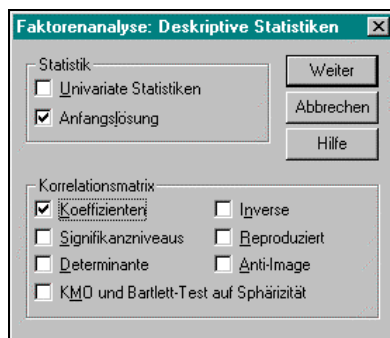


Abbildung 26.19: Dialogfeld der Schaltfläche „Deskriptive Statistik“

Statistiken

- **Univariate Statistiken:** Hiermit fordern Sie eine Tabelle an, in der für jede Variable die Anzahl gültiger Fälle, der Mittelwert und die Standardabweichung angegeben wird.
- **Anfangslösung:** Hiermit werden die Kommunalitäten, die Eigenwerte und der Anteil der erklärten Varianz für die Anfangslösung ausgegeben. Für diese An-

gaben wird keine eigenständige Tabelle erstellt, sondern sie werden in die beiden Tabellen *Kommunalitäten* und *Erklärte Gesamtvarianz* eingefügt.

Korrelationsmatrix

- **Koeffizienten:** Matrix mit den Korrelationen zwischen allen in die Faktorenanalyse einbezogenen Variablen
- **Signifikanzniveaus:** Matrix mit den einseitigen Signifikanzen der Koeffizienten aus der Korrelationsmatrix
- **Determinante:** Gibt die Determinante der Korrelationsmatrix an.
- **KMO und Bartlett-Test aus Sphärizität:** Hiermit werden das Kaiser-Meyer-Olkin-Maß der Stichprobenadäquanz und der Bartlett-Test auf Sphärizität ausgegeben.
- **Inverse:** Gibt die Inverse der Korrelationsmatrix aus.
- **Reproduziert:** Matrix der reproduzierten Korrelationskoeffizienten und deren Residuen
- **Anti-Image:** Gibt eine Matrix mit den Anti-Image-Kovarianzen und eine Matrix mit den Anti-Image-Korrelationskoeffizienten aus. Auf der Hauptdiagonalen der Matrix mit den Anti-Image-Korrelationskoeffizienten wird für jede Variable das Maß der Stichprobenadäquanz (MSA) ausgegeben.

26.8.3 Extraktionsmethode

Abbildung 26.20 zeigt das Dialogfeld der Schaltfläche *Extraktion*, in dem Sie Einstellungen zum Extraktionsverfahren vornehmen und Output zur Beschreibung des Extraktionsergebnisses anfordern können.

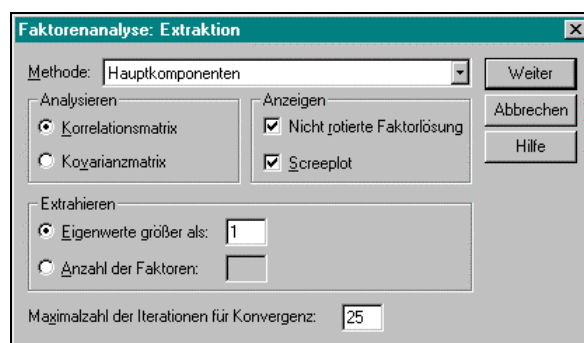


Abbildung 26.20: Dialogfeld der Schaltfläche „Extraktion“

Methode

Wählen Sie in der Dropdown-Liste eines der sieben zur Verfügung stehenden Extraktionsverfahren aus. Das voreingestellte Verfahren *Hauptkomponenten* wurde in dem Beispiel dieses Kapitels verwendet.

Analysieren

Wenn Sie zur Faktorextraktion eine der Methoden *Hauptkomponenten*, *Hauptachsen-Faktorenanalyse* und *Image-Faktorisierung* verwenden, können Sie wählen, ob diese auf einer Analyse der *Korrelationsmatrix* oder der *Kovarianzmatrix* basieren soll. Wenn Sie in der Faktorenanalyse Variablen verwenden, die auf unterschiedlichen Skalen gemessen werden, sollten Sie die Option *Korrelationsmatrix* wählen.

Extrahieren

- **Eigenwerte größer als n:** Als Voreinstellung werden Faktoren mit Eigenwerten > 1 gezogen. Geben Sie eine Zahl zwischen 0 und der Anzahl der in die Analyse einbezogenen Variablen ein, um eine andere Grenze festzulegen.
- **Anzahl der Faktoren:** Alternativ können Sie auch die Anzahl der zu ziehenden Faktoren vorgeben. Es wird dann unabhängig von den Eigenwerten der Faktoren diese Faktorenzahl gezogen. Geben Sie hierzu einen positiven ganzzahligen Wert in das Eingabefeld ein.

Siehe zur Bedeutung dieser beiden Einstellungen auch den Abschnitt *Bestimmung der Faktorenzahl*, S. 650.

Anzeigen

- **Nicht rotierte Fakturlösungen:** Kreuzen Sie diese Option an, damit im Output der Faktorenanalyse eine Tabelle ausgegeben wird, die die Faktorladungen, Kommunalitäten und Eigenwerte für die unrotierte Fakturlösung angibt.
- **Screeplot:** Hiermit wird ein Diagramm erstellt, das die Eigenwerte der Faktoren in absteigender Folge darstellt (siehe S. 651). Wird eine Rotation durchgeführt, stellt das Diagramm die Werte der rotierten Faktoren dar.

Maximalzahl der Iterationen für Konvergenz

Hier können Sie für das Extraktionsverfahren eine Höchstzahl zu durchlaufender Iterationsschritte vorgeben.

26.8.4 Rotation

Das Dialogfeld der Schaltfläche *Rotation* wird in Abbildung 26.21 wiedergegeben. In diesem können Sie festlegen, ob eine Rotation durchgeführt und ggf. welcher Art diese sein soll. Zusätzlich können Sie eine spezielle Tabelle sowie eine Grafik zur Beschreibung der Rotation anfordern.

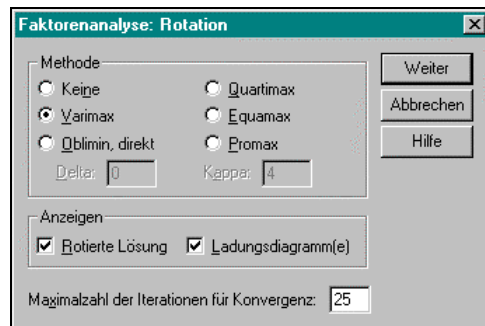


Abbildung 26.21: Dialogfeld der Schaltfläche „Rotation“

Methode

Eine Rotation wird nur durchgeführt, wenn mindestens zwei Faktoren gezogen werden. Wählen Sie für diesen Fall eine der folgenden Optionen:

- **Keine:** Dies ist die Voreinstellung, mit der keine Rotation durchgeführt wird.
- **Varimax:** Es wird eine orthogonale Rotation nach der Varimax-Methode durchgeführt.
- **Oblimin, direkt:** Hiermit wird eine schiefwinklige Rotation nach der Oblimin-Methode durchgeführt. Sie können einen Wert für *Delta* vorgeben, der das Ausmaß bestimmt, in dem die Faktoren miteinander korreliert sein dürfen. Delta muß kleiner oder gleich 0,8 sein. Zu empfehlen sind Werte ≤ 0 . Ein Delta von 0 - dies ist die Voreinstellung - führt zu den am stärksten korrelierten Faktoren. Bei Werten unter 0 nimmt der Grad der Korrelation mit steigendem Betrag des Wertes ab.
- **Quartimax:** Es wird eine orthogonale Rotation nach der Quartimax-Methode durchgeführt.
- **Equamax:** Es wird eine orthogonale Rotation nach der Equamax-Methode durchgeführt.
- **Promax:** Hiermit wird eine schiefwinklige Rotation nach der Promax-Methode durchgeführt, die sich aufgrund des im Vergleich zur Oblimin-Methode geringeren Rechenaufwands insbesondere für große Stichproben empfiehlt. Sie können für *Kappa* einen Wert zwischen 1 und 999 vorgeben. Voreingestellt ist der Wert 4, der im allgemeinen auch nicht geändert werden muß.

Anzeigen

- **Rotierte Lösung:** Wenn Sie eine orthogonale Rotation durchführen, werden mit dieser Option die rotierte Faktorladungsmatrix sowie die Transformationsmatrix ausgegeben. Bei einer schiefwinkligen Rotation (*Oblimin, direkt* oder *Promax*) werden die rotierte Faktorladungs-, Faktorstruktur- und Faktor-korrelationsmatrix ausgegeben. Wenn Sie keine Rotation durchführen, steht diese Option nicht zur Verfügung.
- **Ladungsdiagramm(e):** Führt die Faktorextraktion zu zwei Faktoren, wird ein zweidimensionales Faktorladungsplot erstellt. Im Falle von mehr als zwei Faktoren wird für die ersten drei Faktoren ein dreidimensionales Ladungsplot ausgegeben.

Maximalzahl der Iterationen für Konvergenz

Hier können Sie für die Rotation eine Höchstzahl zu durchlaufender Iterationsschritte vorgeben.

26.8.5 Faktorwerte

Mit der Schaltfläche *Werte* öffnen Sie das in Abbildung 26.22 dargestellte Dialogfeld. Mit den Optionen dieses Dialogfeldes können Sie Variablen mit den Faktorwerten in die Datendatei einfügen.

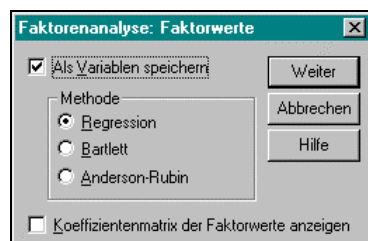


Abbildung 26.22: Dialogfeld der Schaltfläche „Werte“

Als Variablen speichern

Kreuzen Sie diese Option an, um Faktorwerte für die Faktoren berechnen bzw. schätzen und als Variablen in der Datendatei speichern zu lassen. SPSS fügt die neuen Variablen ans Ende der aktiven Datendatei an und weist ihnen automatisch Namen und Labels zu. In die Ausgabedatei wird eine Übersicht mit der Angabe von Namen und Labels der neu erstellten Variablen geschrieben.

Methode

Wenn Sie die Option *Als Variablen speichern* angekreuzt haben, können Sie zum Schätzen der Faktorwerte zwischen den Methoden *Regression*, *Bartlett* und *Anderson-Rubin* wählen. Siehe hierzu auch den Abschnitt *Schätzmethoden*, S. 661.

Koeffizientenmatrix der Faktorwerte anzeigen

Mit dieser Option wird eine Matrix mit den Koeffizienten ausgegeben, mit denen die Variablen multipliziert werden müssen, um Faktorwerte zu erhalten. Zusätzlich wird eine Kovarianzmatrix der Faktorwerte erstellt.

26.8.6 Optionen

In dem Dialogfeld der Schaltfläche *Optionen* aus Abbildung 26.23 können Sie die Behandlung von Fällen mit fehlenden Werten festlegen und die Darstellung von Koeffizienten im Output der Faktorenanalyse steuern.

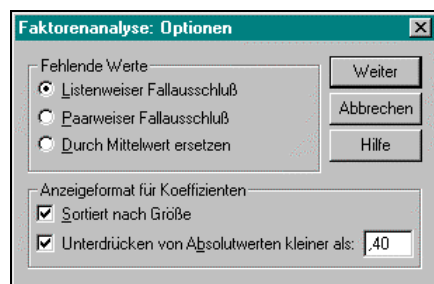


Abbildung 26.23: Dialogfeld der Schaltfläche „Optionen“

Fehlende Werte

Wählen Sie für den Umgang mit Fällen, in denen fehlende Werte enthalten sind, zwischen den folgenden Optionen:

- **Listenweiser Fallausschluß:** Fälle, die in mindestens einer der in die Faktorenanalyse einbezogenen Variablen einen fehlenden Wert aufweisen, werden vollständig aus der Analyse ausgeschlossen.
- **Paarweiser Fallausschluß:** Bei der Berechnung der Korrelationen werden alle Fälle berücksichtigt, die für das betreffende Variablenpaar zwei gültige Werte aufweisen, auch wenn diese Fälle in anderen in die Analyse einbezogenen Variablen einen fehlenden Wert enthalten.
- **Durch Mittelwert ersetzen:** Fehlende Werte werden durch den Mittelwert der jeweiligen Variablen ersetzt, so daß alle Fälle in die Analyse einbezogen werden. Das Ersetzen der fehlenden Werte wirkt sich nur auf die aktuelle Faktorenanalyse und nicht auf die Datendatei aus.

Anzeigeformat für Koeffizienten

Zu empfehlen ist eine Kombination der beiden folgenden Optionen:

- **Sortiert nach Größe:** Faktorladungs- und Faktorstrukturmatrizen werden nach der Größe der Werte sortiert, so daß Variablen, die auf denselben Faktor hohe Ladungen aufweisen, zusammen erscheinen.
- **Unterdrücken von Absolutwerten kleiner als:** Diese Option schließt Koeffizienten mit einem Betrag unterhalb des angegebenen Wertes aus den Darstellungen aus. Damit werden nur die besonders relevanten Werte aufgeführt.

Zur Wirkung dieser beiden Optionen siehe auch *Sortierte Darstellung der Faktorladungen*, S. 659.