Faktorenanalyse

Kurt Holm

Almo Statistik-System
www.almo-statistik.de
kurt.holm@jku.at

2022

Autor: em. Prof. Dr. Kurt Holm, Universität Linz

Im Text wird häufig auf das Dokument **P0** Bezug genommen. Dabei handelt es sich um das Almo-Dokument "Arbeiten mit Almo.PDF" (Dokument 0). Siehe insbesondere auch die PDF-Dokumente "Konfirmatorische Faktorenanalyse" (Dokument 16)
"Bootstrap bei Faktorenanalyse“ (Dokument15a)
"Allgemeine multiple Korrespondenzanalyse" (Dokument 6)

Weitere Almo-Dokumente

Die folgenden Dokumente können alle von der Handbuchseite in www.almo-statistik.de heruntergeladen werden

```
0.  Arbeiten_mit_Almo.PDF (1 MB)
1a. Eindimensionale Tabellierung.PDF  (1.8 MB)
1b. Zwei- und drei-dimensionale Tabellierung.PDF (1.1 MB)
2.  Beliebig-dimensionale Tabellierung.PDF (1.7 MB)
3.  Nicht-parametrische Verfahren.PDF (0.9 MB)
4.  Kanonische Analysen.PDF (1.8 MB)
    enthält: Kanonische Korrelation, Diskriminanzanalyse, bivariate
    Korrespondenzanalyse, optimale Skalierung
5.  Korrelation.PDF (1.4 MB)
6.  Allgemeine multiple Korrespondenzanalyse.PDF (1.5 MB)
7.  Allgemeines ordinales Rasch-Modell.PDF (0.6 MB)
7a. Wie man mit Almo ein Rasch-Modell rechnet.PDF (0.2 MB)
8.  Tests auf Mittelwertsdifferenz,  t-Test.PDF (1,6 MB)
9.  Logitanalyse.pdf  (1,2MB)  enthält Logit- und Probitanalyse
9a. Bootstrap bei Logit- und Probitanalyse.pdf (0,8 MB)
10. Koeffizienten der Logitanalyse.PDF  (0,06 MB)
11. Daten-Fusion.PDF  (1,1 MB)
12. Daten-Imputation.PDF  (1,3 MB)
13. ALM Allgemeines Lineares Modell.PDF  (2.3 MB)
13a. ALM Allgemeines Lineares Modell II.PDF  (2.7 MB)
13b. Bootstrap bei Allgemeinem Linearen Modell.pdf (1 MB)
14. Ereignisanalyse: Sterbetafel-Methode, Kaplan-Meier-Schätzer,
    Cox-Regression.PDF (1,5 MB)
15. Faktorenanalyse.PDF (1,6 MB)
15a. Bootstrap bei Faktorenanalyse.PDF (1,7 MB)
16. Konfirmatorische Faktorenanalyse.PDF (0,3 MB)
17. Clusteranalyse.PDF (3 MB)
18. Pisa 2012 Almo-Daten und Analyse-Programme.PDF (17 KB)
19. Guttman- und Mokken-Skalierung.PFD (0.8 MB)
20. Latent Structure Analysis.PDF (1 MB)
21. Statistische Algorithmen in C (80 KB)
22. Conjoint-Analyse (PDF 0,8 MB)
23. Ausreisser entdecken (PDF 170 KB)
24. Statistische Datenanalyse Teil I, Data Mining I
25. Statistische Datenanalyse Teil II, Data Mining II
26. Statistische Datenanalyse Teil III, Arbeiten mit Almo-
    Datenanalyse-System
27. Mehrfachantworten, Tabellierung von Fragen mit Mehrfach-
    antworten (0.8 MB)
28. Metrische multidimensionale Skalierung (MDS)  (0,4 MB)
29. Metrisches multidimensionales Unfolding (MDU) (0,6 MB)
30. Nicht-metrische multidimensionale Skalierung (MDS)  (0,4 MB)
31. Pfadanalyse.PDF  (0,7 MB)
32. Datei-Operationen mit Almo  (1,1 MB)
33. Wählerstromanalyse und Wahlhochrechnung (1,6 MB)
34. Soziometrie. Auswertung soziometrischer Daten (0,5 MB)
35. Konfidenzintervall und p-Wert beim Bootstrap-Verfahren (200 KB)
```

INHALTSVERZEICHNIS

P30 Faktorenanalyse

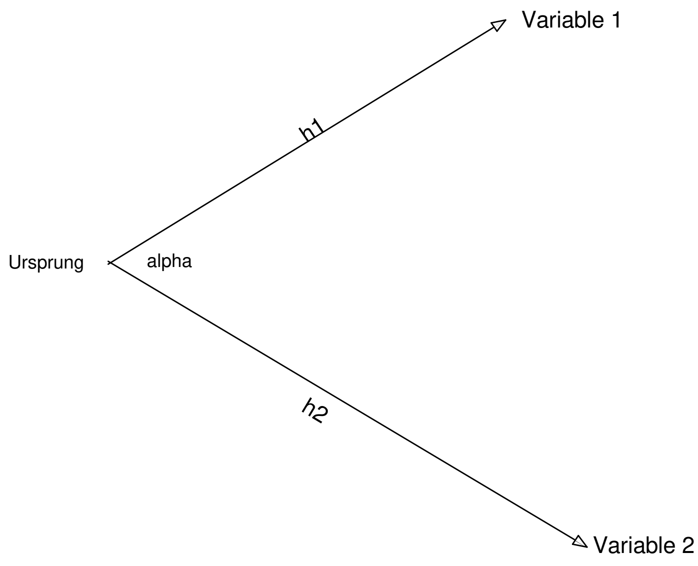
Programm 30 rechnet eine Faktorenanalyse mit rechtwinkeliger oder schiefwinkeliger Rotation für quantitative, ordinale und nominale Variable. Das Programm wurde von Kurt Holm geschrieben. Die schiefwinkelige Rotation verwendet das Quartimin-Verfahren mit nachträglicher Verbesserung durch das Gruppenrotationsverfahren. Die rechtwinkelige Rotation verwendet das Varimax-Verfahren.

Im vorliegenden Handbuch wird ausschließlich die quantitative Faktorenanalyse behandelt. Eine echte ordinale Faktorenanalyse wird in Almo nicht angeboten. Ordinale Variable werden durch den "Groß-Gamma-Kalkül" in eine Korrelationsmatrix überführt, die dann mit der quantitativen Faktorenanalyse faktorisiert wird. Siehe dazu das Handbuch 5 "Korrelation". Die nominale Faktorenanalyse wird im Handbuch Nr. 6 "Korrespondenzanalyse - nominale Faktorenanalyse" behandelt.

P30.1 Einführung

Die Ausführungen in diesem Abschnitt sind teilweise entnommen (dabei überarbeitet und verändert) aus Holm (Die Befragung 3, UTB 433).

Die Ausgangsdaten für eine Faktorenanalyse bestehen üblicherweise aus einer Korrelationsmatrix. Wir wollen annehmen, wir hätten Meßwerte von 4 Variablen. Diese 4 Variablen werden interkorreliert. Wir erhalten eine 4*4 Korrelationsmatrix.
In der Hauptdiagonalen dieser Matrix stehen die Eigenkorrelationen, d. h. die Korrelationen der Variablen mit sich selbst. Diese sind selbstverständlich 1.0.
Wir müssen zuerst erklären, was die Korrelation zwischen zwei Variablen geometrisch betrachtet ist.



In der Faktorenanalyse betrachten wir eine Variable als einen Vektor, der die Länge h besitzt. Die Korrelation zwischen zwei Variablen wird dann durch den Winkel α zwischen den sie repräsentierenden Vektoren und durch deren Länge ausgedrückt. Je kleiner der Winkel (bei konstanten Längen), um so größer ist die Korrelation. Präzise gilt

$$(1) r_{12} = \cos\alpha \cdot h_1 \cdot h_2$$

Wenn eine 3. und 4. Variable mit hinzugenommen wird, dann läßt sich das in aller Regel nicht mehr in der zweidimensionalen Fläche graphisch darstellen. Bei drei Variablen entsteht dann ein Körper, bei vier Variablen und mehr dann sozusagen ein "Hyper-Körper".

P30.1.1 Kommunalitäten

Wenn wir Variable 1 mit sich selbst korrelieren, entsteht selbstverständlich ein Korrelationskoeffizient von r = 1.0. Geometrisch bedeutet dies, dass der Winkel $\alpha = 0$, d. h. $\cos\alpha = 1.0$ wird. Setzen wir nun in obige Gleichung 1 entsprechende Werte ein, dann entsteht jedoch folgendes

$$(2) r_{11} = \cos 0^\circ \cdot h_1 \cdot h_1 = h_1^2$$

Die EigenKorrelation im faktorenanalytischen Modell ist also nicht gleich 1.0, sondern gleich h^2.
Dieses h^2 wird in der Faktorenanalyse "Kommunalität" genannt. In der Korrelationsmatrix dürfen also in der Hauptdiagonalen, wo ja die Eigenkorrelationen stehen, keine 1.0 stehen. Dort müssen die Kommunalitäten der jeweiligen Variablen stehen.

Nicht selten jedoch wird eine Faktorenanalyse mit 1.0 in der Hauptdiagonalen der Korrelationsmatrix durchgeführt. Es ist üblich, diese besondere Art der Faktorenanalyse dann "Hauptkomponenten-Analyse" zu nennen. Geometrisch gesehen wird hier unterstellt, dass die Vektoren alle von der Länge h = 1 sind. Die Endpunkte der Vektoren (die die Variablen repräsentieren) liegen also bei einem 2-dimensionalen System auf einem Radius von l um den Ursprung, bei einem 3-dimensionalen System auf der Oberfläche einer Kugel vom Radius 1 um den Ursprung und bei einem mehrdimensionalen System auf der Oberfläche einer Hyperkugel vom Radius l um den Ursprung.

P30.1.2 Faktorisierung

Wir wollen nun eine Interkorrelationsmatrix von 4 Variablen betrachten. Wir wollen annehmen, wir hätten 4 sportliche Leistungen an 100 Personen gemessen und dabei folgende Korrelationsmatrix gefunden.

Korrelationsmatrix 1

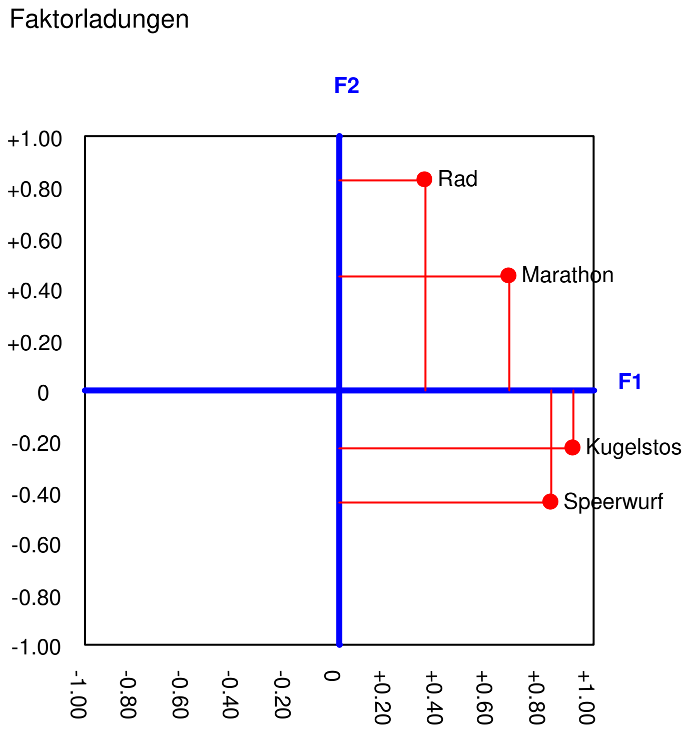
		Marathon V1	Rad V2	Kugel V3	Speer V4
Marathon	V1	0.65	0.60	0.51	0.36
Rad	V2	0.60	0.80	0.12	-0.08
Kugelstoss	V3	0.51	0.12	0.90	0.57
Speerwurf	V4	0.36	-0.08	0.87	0.89

In der Hauptdiagonalen dieser Matrix haben wir nicht 1.0, sondern die Kommunalitäten h^2 eingetragen. Diese kennen wir, weil wir das Beispiel "rückwärts" konstruiert haben. Wir haben zuerst die unten angegebene Faktorladungsmatrix **L** angenommen, aus der wir dann die Korrelationsmatrix gemäß der Formel **R=L*L'** rekonstruiert haben.

Wenn wir aus den h^2 -Werten die Wurzel ziehen, dann erhalten wir die Länge h des Vektors der vier Variablen

0,806 0,894 0,949 0,943

Die geometrische Abbildung dieser Korrelationsmatrix ist nun folgende



Die Koordinaten der vier Punkte im Koordinatensystem sind

```
Matrix der Faktorladungen
```

		Faktor 1	Faktor 2
Marathon	V1	0.6692	0.4495
Rad	V2	0.3410	0.8268
Kugelstoss	V3	0.9193	-0.2341
Speerwurf	V4	0.8340	-0.4408

Diese Matrix der Koordinaten wird in der Faktorenanalyse "Faktorladungsmatrix" genannt. Wenn wir in der Zeichnung 2 die Länge h; der Vektoren der vier Variablen nachmessen (oder aus den uns bekannten Koordinaten ausrechnen), dann erhalten wir

0,806 0,894 0,949 0,943

Werden diese Werte quadriert, dann erhalten wir die in der Diagonalen der Korrelationsmatrix eingetragenen Kommunalitäten h^2.
Die Koordinatenachsen F1 und F1 werden "Faktoren" genannt. Die Faktorenanalyse hat also ergeben, dass "hinter" den vier Variablen zwei Faktoren stehen. Wir erkennen aus Zeichnung 2, dass die Variablen 1, 2 und die Variablen 3, 4 je eine "Punktwolke" bilden. Die Faktorenanalyse hat uns also nachgewiesen, dass die vier Variablen nicht auf *einer* gemeinsamen Dimension messen. Das ist bei diesem einfachen Beispiel auch bereits an den Korrelationen erkenntlich. Variable 1 und 2 korrelieren mit 0,6. Variable *3* und 4 korrelieren mit 0,87. Die "Quer"-Korrelationen der beiden Gruppen sind mit 0.51, 0.36, 0.12, -0,08 deutlich geringer.
Wir wollen nun unter Verwendung von Gleichung (1) überprüfen, ob Zeichnung 2 tatsächlich eine geometrische Darstellung der Korrelationsmatrix ist.
Betrachten wir beispielsweise Variable 2 und 3. Der Winkel zwischen diesen beiden Vektoren beträgt, wie wir durch Nachmessen ermitteln, $\alpha = 81° \; 50'$. Dann ist $\cos \alpha = 0,14$. Wir setzen nun in Gleichung l ein und erhalten

$$\begin{aligned} r_{23} &= \cos \alpha * h_2 * h_3 \\ &= 0,14 * 0,894 * 0,949 \\ &= 0,12 \end{aligned}$$

Der Leser kann nun nach diesem Schema auch die übrigen Korrelationskoeffizienten reproduzieren.
Er wird feststellen, dass Zeichnung 2 eine adäquate geometrische Darstellung der Korrelationsmatrix ist.
Wir symbolisieren die Korrelationsmatrix mit den Kommunalitäten h^2 in der Diagonalen durch $\mathbf{R_h}$ und die Faktorladungsmatrix mit **L**. Die Transponierte von **L** ist dann **L'**.
Die Transponierte **L'** entsteht sehr einfach dadurch, dass wir bei der Faktorladungsmatrix die Faktoren F1 und F1 als Zeilen und die Variablen l bis 4 als Spalten schreiben.
Es gilt nun

$$(3) \qquad \boldsymbol{R_h} = \boldsymbol{L} \cdot \boldsymbol{L'}$$

Diese Gleichung wird das "Fundamentaltheorem" der Faktorenanalyse genannt. Gleichung 3 läßt sich auch in folgender Weise schreiben

$$(3a) \quad R_h = R_{1.0} - U^2 = L \bullet L'$$

$\boldsymbol{R_{1.0}}$ = Korrelationsmatrix mit 1.0 in der Diagonalen
$\boldsymbol{U^2}$ = Diagonalmatrix. In ihrer Diagonalen steht $1.0 - h_i^2$;
außerhalb der Diagonalen steht .0
h_i^2 sind die Kommunalitäten der Variablen i.

Der Leser kann die in (3) beschriebene Multiplikation durchführen. Er wird erkennen, dass wenn **L** mit der transponierten **L'** postmultipliziert wird, die Korrelationsmatrix entsteht.
Wir sind nun bereits in der Lage, die l. und zweifellos auch wichtigste Aufgabe, die durch eine Faktorenanalyse gelöst werden soll, zu definieren:

Die Faktorenanalyse soll eine Korrelationsmatrix $\mathbf{R_h}$ in eine Faktorladungsmatrix **L** überführen – geometrisch gesprochen: sie soll eine Korrelationsmatrix $\mathbf{R_h}$ als einen Punkteschwarm darstellen, der in ein rechtwinkeliges Koordinatensystem eingebettet ist.
Das ist die Aufgabe der "Faktorisierung" einer Korrelationsmatrix. Diese Aufgabe wird durch das mathematische "Eigenwert-Verfahren" gelöst, das wir hier nicht darstellen wollen.

Betrachten wir noch einmal Zeichnung 2. Wir können das System der beiden Koordinatenachsen F1 und F1 um den Ursprung in eine andere Position drehen – dabei die Punktwolke jedoch stehen lassen. Die Koordinaten der vier Punkte, also die L-Matrix würde sich dabei ändern. Das geometrische Bild der Punktwolke, vor allem die Abstände der vier Punkte zueinander, würde erhalten bleiben. Das bedeutet, dass die Lage des Koordinatensystems relativ gleichgültig ist.

Wenn wir nun auf die Korrelationsmatrix das oben erwähnte "Eigenwert-Verfahren" anwenden, dann wird die Achse F1 so durch die Punktwolke gelegt, dass die auf sie projizierten Punkte, also die Koordinaten, quadriert, sich zu einem Maximum aufsummieren. Statistisch bedeutet dies, dass die Achse F1 eine varianzmaximierende Lage besitzt. Die Achse F1 wird dann orthogonal zur Achse F1 gelegt.
Wir wollen jedoch nochmals betonen, dass die jeweilige Lage des Koordinatensystems beliebig ist. Es kommt, wenn wir inhaltliche Interpretation leisten wollen, nur auf die Gestalt der Punktwolke an.

Noch einmal: Kommunalitätenschätzung
Bei empirischen Daten kennen wir natürlich nicht die Kommunalitäten. Wir müssen sie schätzen.
Wir rechnen eine 1. Analyse, bei der wir in die Diagonale der Korrelationsmatrix die R_i^2 eingesetzt. R_i^2 ist der quadrierte multiple Korrelationskoeffizient der Variablen i mit den anderen Variablen. Danach werden 20 Kommunalitäten-Iterationen durchgeführt. Wir werden später, in Abschnitt P30.2.6, diesen Vorgang erläutern.

Dann rechnen wir eine 2. Analyse, wobei wir in die Diagonale 1.0 einsetzen.
Almo ermittelt schließlich folgende Faktorladungen.

Matrix der Faktorladungen

	richtig		mit R-Quadrat und 20 Iterationen		mit 1.0	
	Faktor 1	Faktor 2	Faktor 1	Faktor 2	Faktor 1	Faktor 2
Marath	0.6692	0.4495	0.6900	0.5052	0.7848	0.4771
Rad	0.3410	0.8268	0.3142	0.7579	0.4037	0.8563
Kugel	0.9193	-0.2341	0.9023	-0.2183	0.9098	-0.3219
Speer	0.8340	-0.4408	0.8533	-0.4568	0.8172	-0.5229

Der Unterschied zu den richtigen Faktorladungen ist gering. Bei der grafischen Darstellung der drei Faktorladungsmatrix erkennt man praktisch dieselbe Struktur der Punkte.

P30.1.3 Faktorenzahl

Unser Zahlenbeispiel ist konstruiert. Die vier Variablen liegen exakt in einem 2-dimensionalen Raum. Bei empirischen Zahlen wäre das nicht der Fall.

Wir wollen die Korrelationsmatrix nur ganz wenig verändern. Wir schreiben die Kommunalitäten gleich in die Diagonale.

Korrelationsmatrix 2

		Marathon V1	Rad V2	Kugel V3	Speer V4
Marathon	V1	0.6600	0.6100	0.5000	0.3500
Rad	V2	0.6100	0.8100	0.1100	-0.0800
Kugelstoss	V3	0.5000	0.1100	0.9100	0.8800
Speerwurf	V4	0.3500	-0.0800	0.8800	0.9000

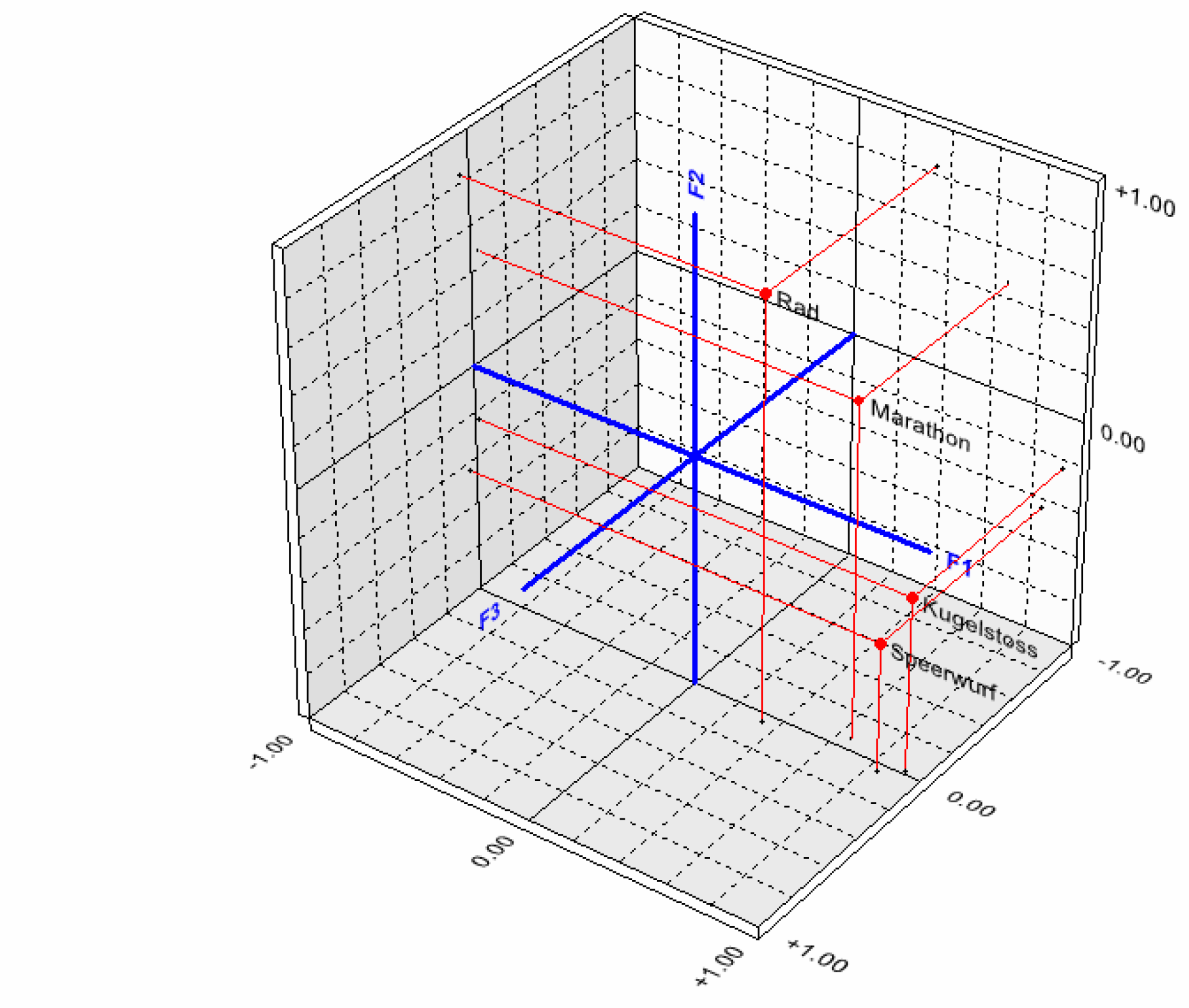
Es ist deutlich zu erkennen, dass wir nur + 0,01 bzw. – 0,01 hinzugefügt haben.
Die Faktorenanalyse liefert uns nun folgende L-Matrix

Matrix der Faktorladungen

		Faktor 1	Faktor 2	Faktor 3
Marathon	V1	0.6626	0.4669	-0.0525
Rad	V2	0.3372	0.8331	0.0464
Kugelstoss	V3	0.9228	-0.2405	-0.0173
Speerwurf	V4	0.8402	-0.4383	0.0418

Zur geometrischen Darstellung benötigen wir nun ein 3-dimensionales Koordinatensystem.

Faktorladungen

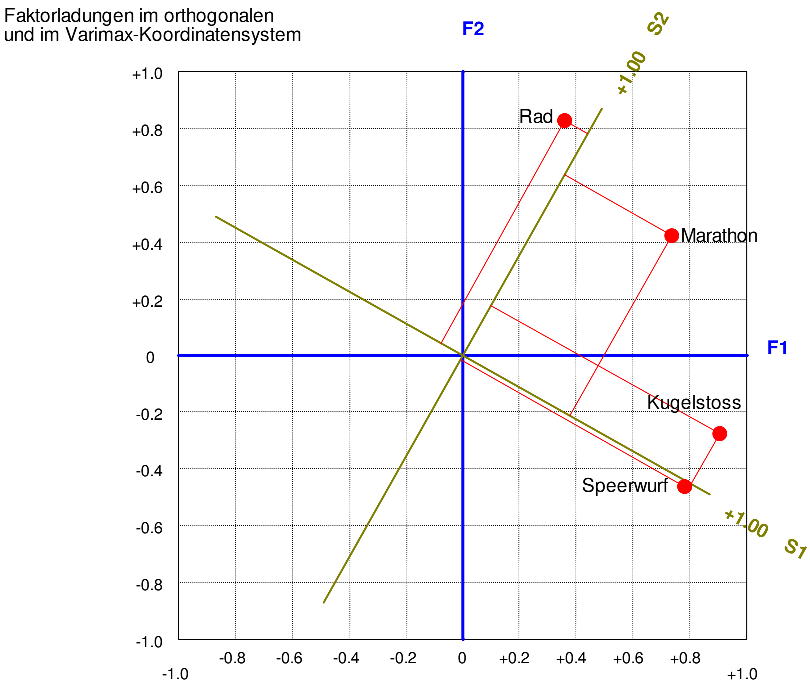


Die vier Punkte weichen aus der durch F1-F2 gebildeten Ebene nur minimal in die 3. Dimension F3 ab.
Die Frage ist, ob diese Abweichungen lediglich Zufallsabweichungen von einer tatsächlich nur 2-dimensionalen Struktur sind. Wir müssen ja berücksichtigen, dass unsere Korrelationskoeffizienten nicht "wahre" Populations-, sondern fehlerbehaftete Stich-probenwerte sind.
Anders formuliert: Die Frage ist, ob zwei Faktoren genügen, die Korrelationsmatrix befriedigend zu reproduzieren.
Das ist die 3. Aufgabe, die innerhalb der Faktorenanalyse gestellt ist, die Aufgabe, die Zahl der signifikanten Faktoren zu bestimmen.
Wenn wir für die obige (zweite etwas veränderte) Korrelationsmatrix die Faktoren F1 und F2 als ausreichend betrachten und gemäß Gleichung 3 die Korrelationsmatrix reproduzieren, dann werden wir feststellen, dass die reproduzierten Korrelationen in den einzelnen Zellen nur um Werte von 0,01 von den tatsächlichen Korrelationen abweichen. Das scheint so minimal zu sein, dass wir behaupten dürfen, dass zwei Faktoren die Korre1ationsmatrix befriedigend darstellen.

P30.1.4 Rotation

Nun wollen wir nochmals die Grafik in P30.1.2 betrachten. Wir sagten, dass die Faktorenanalyse uns die räumliche Struktur der 4-Punkte-Wolke enthüllt – wobei die Lage der Koordinatenachsen F1 und F2 beliebig sind. Das mathematische Verfahren der "Eigenwert"-Berechnung zeigt uns die Gestalt der Punktewolke. Die Lage der Punktewolke im Koordinatensystem ist jedoch nur eine von vielen möglichen. Wenn es darum geht, das Ergebnis der Faktorenanalyse inhaltlich zu erklären, dann ist dies sogar eine nicht brauchbare Lage.

Die rechtwinklige Rotation
Die Punktewolke muss im Koordinatensystem in eine interpretierbare Lage gedreht (rotiert) werden - oder (was das gleiche ist) das Koordinatensystem wird gedreht. Das ist dann die einfachste "Rotationsmethode": die rechtwinklige Rotation. Dazu wird üblicherweise das *Varimax*-Verfahren verwendet. Für unser Beispiel entsteht folgende Grafik:



Die sandfarbige Achse S1 kann nun als der Faktor der Kraft interpretiert werden und S2 als Faktor der Ausdauer. Werden die 4 Punkte auf das neue sandfarbene Koordinatensystem projiziert dann erhalten wir folgende Faktorladungsmatrix:

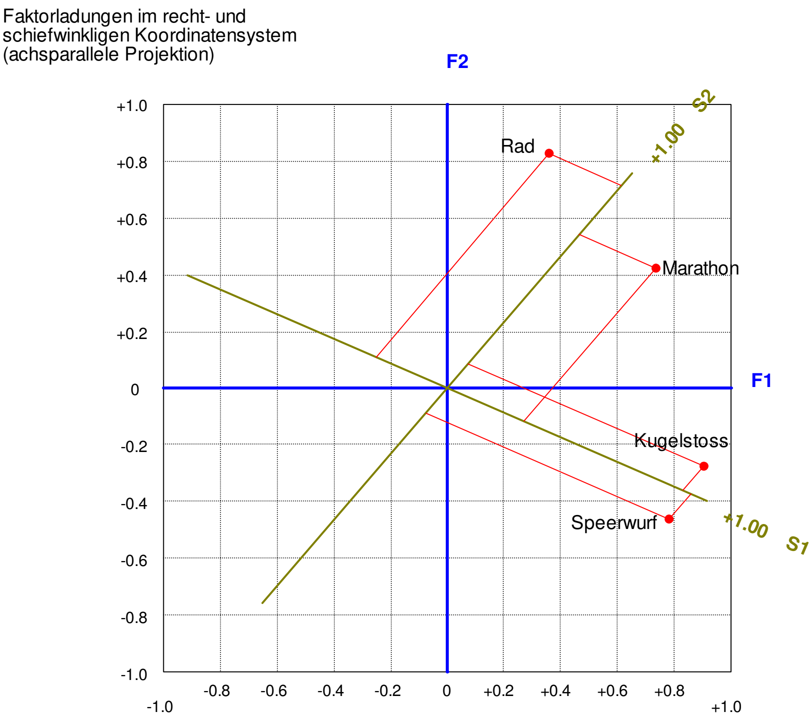
```
Rechtwinklig Varimax-rotierte Faktorladungen
```

		Faktor 1	Faktor 2
Marathon	V1	0.4346	0.7323
Rad	V2	-0.0908	0.8986
Kugelsto	V3	0.9244	0.2053
Speerwur	V4	0.9110	-0.0186

In Abschnitt P30.2.3.7 beschreiben wir die *benutzerdefinierte rechtwinklige Rotation* mit 2 und 3 Faktoren. Bei dieser Methode kann der Benutzer im Grafik-Editor die Punktekonfiguration beliebig rechtwinklig rotieren und sich dann die dazugehörende Faktorladungsmatrix ausgeben lassen.

Die schiefwinklige Rotation

Noch besser interpretierbare Ergebnisse erhalten wir, wenn die Achsen jeweils mitten zwischen den beiden Punkten, die eine "Punktewolke" bilden, hindurch laufen. Dann allerdings werden die beiden Achsen nicht mehr rechtwinklig aufeinander stehe. Das ist die schiefwinklige Rotation, die in Almo nach dem Quartimin-Verfahren (mit anschliessender Verbesserung durch die Gruppenrotation) erfolgt. Es entsteht folgende Grafik



Wir legen "mitten durch" die Gruppe der Variablen 1, 2 die Achse S1 und "mitten durch" die Gruppe der Variablen 3, 4 die Achse S2. Die Achse S1 repräsentiert nun die Meßdimension, auf der die Variablen 1, 2 messen, und die Achse S2 repräsentiert die Meßdimension, auf der die Variablen 3, 4 messen.

Wir können nun die vier Punkte achsparallel (d. h. parallel zu den Achsen S1, S2) auf die neuen Achsen projizieren. Wir erhalten folgende neue Faktorladungsmatrix **K**.

```
Matrix der auf die schiefwinkligen Achsen
```

achsparallel projizierten Faktorladungen
(Ladungsmatrix)

		Faktor 1	Faktor 2
Kugelsto	V3	0.9072	0.1135
Speerwur	V4	0.9390	-0.1181
Marathon	V1	0.2971	0.7152
Rad	V2	-0.2761	0.9453

Der Winkel α zwischen den Achsen S1 und S2, genauer cos α gibt an, wie stark die beiden Meßdimensionen miteinander korrelieren. ALMO liefert uns folgende Ausgabe:

Matrix der Winkel zwischen den schiefwinkligen Achsen

	Faktor 1	Faktor 2
Faktor 1	0	72.8367
Faktor 2	72.8367	0

Matrix der Korrelationen zwischen den schiefwinkligen Achsen

	Faktor 1	Faktor 2
Faktor 1	1.0000	0.2951
Faktor 2	0.2951	1.0000

Das "Ziehen" der Achsen S1 und S2 wird in der faktorenanalytischen Literatur "Rotation" genannt. Die Achsen werden aus der Lage F1, F1 in die neue Lage S1, S2 rotiert.
In unserem Beispiel haben wir so rotiert, dass der rechte Winkel zwischen den beiden Achsen verloren ging. Man nennt das "schiefwinklige Rotation". Wird der rechte Winkel beibehalten, dann wird von "orthogonaler Rotation" gesprochen. Es ist unmittelbar zu erkennen, dass bei orthogonaler Rotation die Achsen S1 und S2 nicht in eine inhaltliche so sinnvolle Lage rotiert werden können.
Obige Faktorladungsmatrix **K** wird üblicherweise "rotierte Faktorladungsmatrix" genannt. Die uns bereits bekannte **L**-Matrix wird entsprechend "unrotierte Faktorladungsmatrix" genannt.
Wir können nun das in Gleichung 3 formulierte "Fundamentaltheorem" der Faktorenanalyse in einer neuen Weise formulieren.

(3b) $\mathbf{R_h}=\mathbf{K}*\mathbf{C}*\mathbf{K}'$

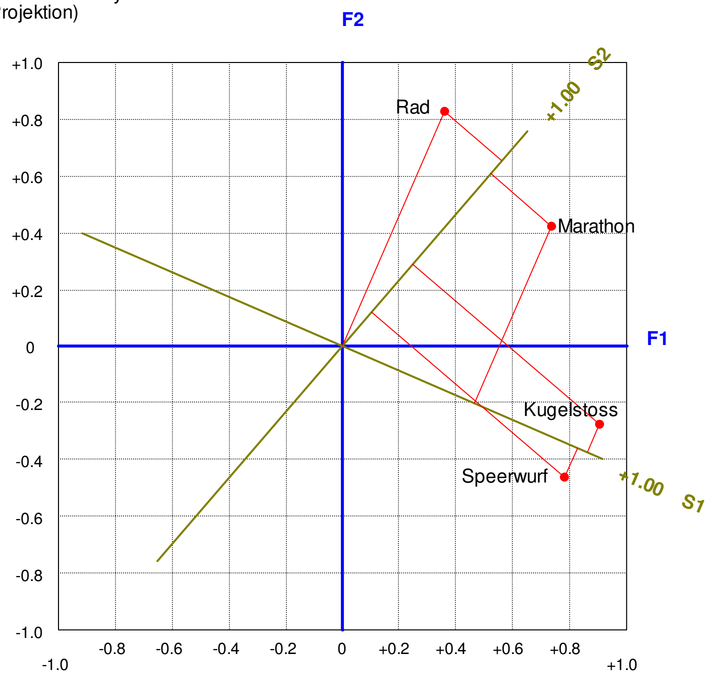
dabei ist **C** die Matrix der Korrelationen zwischen den schiefwinkligen Achsen. Die orthogonale Ladungsmatrix **L**, die wir durch das mathematische "Eigenwert-Verfahren" gewinnen, liefert nur einen Bezugsrahmen für die Matrix **K**, die wir eigentlich suchen.

Wir haben in Zeichnung 3 die vier Punkte "achsparallel" auf die Achsen S1 und S2 projiziert. Wir hätten sie aber auch rechtwinklig auf diese Achsen projizieren können. Die dabei entstehende Matrix von Ladungen (Koordinaten) symbolisieren wir durch **Q**. Harman (1967, Kap. 13) bezeichnet die **Q**-Matrix als "factor-structure" und die **K**-Matrix als "factor-pattern". In der deutschsprachigen Literatur zur Faktorenanalyse ist die Begriffsverwendung nicht einheitlich. SPSS verwendet die Begriffe "Mustermatrix" und "Strukturmatrix" Wir wollen hier, um Verwechslungen zu vermeiden, folgende Bezeichnungen verwenden:

Matrix **K** = **"Ladungsmatrix":**
Matrix der achsparallel projizierten Faktorladungen

Matrix **Q** = **"Strukturmatrix"**:
Matrix der rechtwinklig (oder orthogonal) projizierten Faktorladungen.

Faktorladungen im recht- und schiefwinkligen Koordinatensystem (rechtwinklige Projektion)



In Abschnitt P30.1.9 und dann in P30.2.3.5 werden wir nochmals auf die Rotation zurückkommen.

P30.1.5 Faktorenwerte

Die Faktorladungen auf den neuen Achsen S1, S2 können als Gewichtszahlen verwendet werden, um einen gewichteten Gesamtpunktwert für jede Untersuchungseinheit zu ermitteln.
Auf der Meßdimension S1 liegen die beiden Variablen 1 und 2. Aus diesen beiden Variablen bilden wir nun für jede Untersuchungseinheit einen gewichteten Gesamtpunktwert. Als Gewichtszahlen werden Zahlen verwendet, die lineare Transformationen der orthogonal projizierten Faktorladungen darstellen.
Die Gewichtszahlen werden Faktor-Betaladungen oder Faktorwertkoeffizienten genannt.
Dieser Gesamtpunktwert wird inhaltlich als die bestmögliche Schätzung des Wertes der Untersuchungseinheit auf der Meßdimension S1 begriffen. In der Faktorenanalyse wird dieser Gesamtpunktwert "Faktorwert" genannt. Dies ist nun die 5. und letzte Aufgabe, die innerhalb der Faktorenanalyse zu lösen ist, die Aufgabe der Faktorwertberechnung.

Einige Probleme im Detail

Wir wollen im folgenden ausführlich auf 4 zentrale Probleme der Faktorenanalyse eingehen. Dies sind (1) Zahl der signifikanten Faktoren, (2) Kommunalitätenschätzung, (3) Faktorwerte und (4) Rotation.

P30.1.6 Die Zahl der signifikanten Faktoren

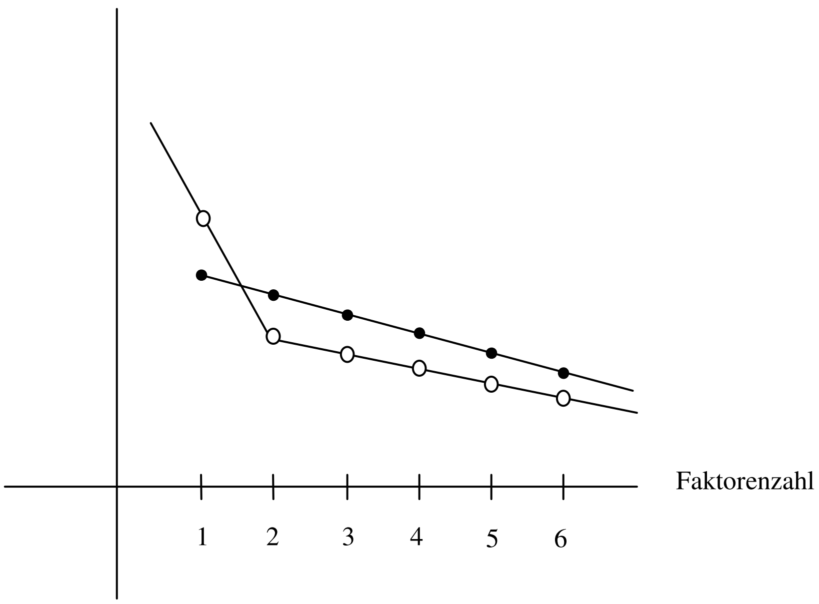
Eine Frage, die die Faktorenanalyse beantworten soll, lautet: Wie viele Faktoren "stecken" in einer Korrelationsmatrix?

Bei entsprechender Wahl der Kommunalitäten, wenn man 1.0 in die Diagonale der Korrelationsmatrix einsetzt, kann man alle Faktoren extrahieren, d. h. man kann so viele Faktoren extrahieren, wie die Matrix Variable umfaßt. Dies gilt auch für Korrelationsmatrizen, die aus Zufallszahlen für m Variable erstellt werden. Die Ladungen der m Variablen auf den m Faktoren werden nicht gleich 0 sein, da die Interkorrelationen zwischen den m Variablen durch Zufallsfehler verzerrt sind. In "echten" empirischen Korrelationsmatrizen stecken die Werte aus einer Stichprobe von N Untersuchungseinheiten und nicht die Werte aus der Grundgesamtheit. Der Zufallsfehler stört die Logik der Matrix. Er ist verantwortlich dafür, dass man (mit 1.0 in der Diagonalen) m Faktoren extrahieren kann.

Was wir suchen, ist ein Verfahren, das die störende Wirkung der Zufallsfehler, die in der Korrelationsmatrix stecken, neutralisiert, ein Verfahren, das uns angibt, ob ein Faktor noch ein echter Faktor ist oder nur durch das Wirken des Zufalls zustande kam. Mit anderen Worten: wir suchen ein Verfahren, das uns hilft, die Zahl der signifikanten Faktoren festzulegen.
In der Literatur wurden viele Methoden schon vorgeschlagen und diskutiert. Hier sollen nur diejenigen vorgetragen werden, die praktische Bedeutung erlangten.

P30.1.6.1 Der Scree-Test

Man kann mit einem Computer sehr leicht normalverteilte und standardisierte Zufallszahlen erzeugen. Werden diese interkorreliert, dann entsteht eine Matrix aus Zufallskorrelationen. Diese wird einer Faktorenanalyse (mit 1.0 in der Diagonalen) unterzogen, wobei m Faktoren (so viel wie Variable) extrahiert werden. Wir haben dies für eine 6 x 6 Matrix getan. (Siehe auch die Darstellung bei Überla (1968, S. 128))



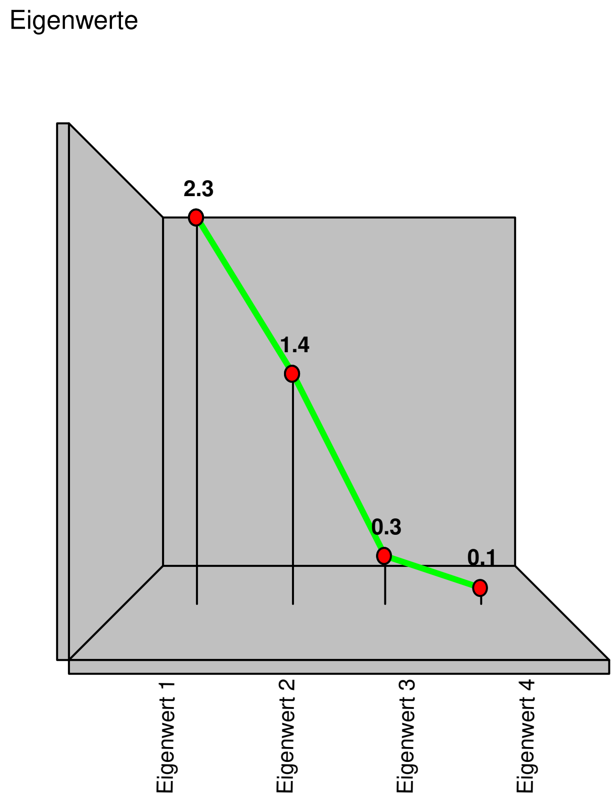
Punkte = Eigenwerte der Zufallskorrelation,
Kreise = Eigenwerte einer empirischen Matrix mit einem signifikanten Faktor.

Die Eigenwerte einer 6 x 6-Zufallskorrelationsmatrix unterscheiden sich voneinander nur geringfügig. Graphisch, in einem Koordinatensystem eingetragen, ergeben sie eine sanft geneigte Kurve. In das Koordinatensystem wurden nun gleichzeitig auch die Eigenwerte für eine empirische 6 x 6-Matrix mit 2 signifikanten Faktoren eingetragen. Es ist deutlich zu erkennen, dass deren 3. und die nachfolgenden Eigenwerte unterhalb der entsprechenden Eigenwerte der Zufallsmatrix liegen. Hingegen liegt der l. und 2. Eigenwert der beiden "echten" Faktoren deutlich über den Eigenwerten des 1. und 2. Faktors aus der Zufallsmatrix.
Wir können daraus folgern, dass der 3. und die weiteren Faktoren unseres Beispiels "Zufallsfaktoren" sind. Die Matrix besitzt zwei signifikante Faktoren.

Wir können als **Regel** formulieren:

Die Folge der signifikanten Faktoren wird da abgebrochen, wo die Kurve der Eigenwerte beginnt flach zu verlaufen.

Wir haben für unser Beispiel mit der 2. Korrelationsmatrix in Abschnitt P30.1.3 (mit 1.0 in der Diagonalen) die Eigenwerte berechnet. ALMO liefert folgende Grafik:



Ab dem 3. Eigenwert verläuft die Kurve flach. Der Scree-Test zeigt uns also 2 signifikante Faktoren an.

P30.1.6.2 Der Test nach Rippe

Bei der Faktorenanalyse wird die Matrix **R** "umgewandelt" in eine Matrix der Faktorladungen L. Wird diese mit ihrer transponierten **L'** multipliziert, dann wird **R** reproduziert. Der Gedanke liegt nun nahe, die Güte dieser Reproduktion zu überprüfen, d. h. nach jeder Faktorextrahierung die (nun um eine Spalten

vergrößerte) Faktorladungsmatrix **L** daraufhin zu untersuchen, wie nahe sie (nach Multiplikation mit ihrer Transponierten) der beobachteten Matrix **R** kommt.
Ein solcher Test wurde zuerst im Rahmen der Maximum-Likelihood-Methode von Lawley und später von Rippe entwickelt. Beide Tests werden bei Harman (1967, S. 219, 197) dargestellt.
Die Gleichung für den Rippe-Test lautet:

$$(1) \text{Chi} - \text{Quadrat} = (\text{N} - 1) \cdot \ln \frac{\det \mathbf{Repro*}}{\det \mathbf{R}}$$

N = Stichprobe (Zahl der Untersuchungseinheiten).
Repro* = das ist die reproduzierte Korrelationsmatrix, die aus der Multiplikation der Faktorladungsmatrix **L** mit ihrer Transponierten entsteht – mit der Besonderheit (angedeutet durch den Sternindex) –, dass in die Hauptdiagonale der reproduzierten Matrix 1.0 eingesetzt wird.
R = das ist die beobachtete Korrelationsmatrix (mit 1.0 in der Diagonalen).
det = Determinante
ln = symbolisiert den natürlichen Logarithmus (zur Basis e).

Anmerkung: Die Determinante der Korrelationsmatrix bezeichnet die "verallgemeinerte Varianz" (engl."generalized variance") der Menge der Variablen $x_1 \ldots x_n$. In Gleichung 1 wird also die durch m Variable erklärte verallgemeinerte Varianz mit der totalen verallgemeinerten Varianz verglichen. Zum Begriff der "verallgemeinerten Varianz" siehe P20.9.4.

Der errechnete Chi-Quadrat-Wert muss mit dem Tabellenwert verglichen werden. Die Zahl der Freiheitsgrade ist:

$$(2) \qquad df = (\frac{1}{2}(m-f)^2 + m - f$$

m = Zahl der Variablen, f = Zahl der Faktoren (mit denen die reproduzierte Korrelationsmatrix erzeugt wird.)
Für unser Beispiel mit der 2. Korrelationsmatrix, wobei wir N=100 unterstellt haben, erhalten wir:

```
Faktoren-Signifikanztest nach Rippe:
-------------------------------------
Chi-Quadrat-Werte je Faktor
   74.5452      0.0000      0.0000      0.0000

Freiheitsgrade
      6            3            1            0

Wahrscheinlichkeit (1-p)*100, dass Faktor  signifikant
   99.9999     0.0000     0.0000      0.0000

Beachte: In die Zahl der signifikanten Faktoren muss der
erste nicht mehr signifikante Faktor miteinbezogen werden
 - sofern ein solcher von Almo noch ausgegeben wird.
Ist schon der 1. nicht signifikant, dann ist dieser signifikant (!)

Determinate der durch die Faktoren reproduzierten Matrix
    0.1781      0.0260      0.0423      0.0839
```

Man beginnt zuerst mit f = 1 Faktor. Die Faktorladungsmatrix besteht also nur aus einer Spalte.
Es wird der Chi-Quadrat- und df-Wert bestimmt. Ist der errechnete Chi-Quadrat-Wert *höher* als der Tabellenwert, dann ist der Faktor signifikant. Danach werden f = 2 Faktoren verwendet.

Die Faktorladungsmatrix besteht dann aus zwei Spalten.
Ist der errechnete Chi-Quadrat-Wert höher als der Tabellenwert, dann ist auch der zweite Faktor signifikant ... usw., bis schließlich ein Chi-Quadrat-Wert errechnet wird, der nicht mehr größer ist als der entsprechende Tabellenwert. Der entsprechende Faktor ist noch signifikant. Alle darauffolgenden Faktoren sind dann nicht mehr signifikant, (sofern die Faktorladungsmatrix spaltenweise nach fallender Größe der Eigenwert angeordnet wurde).
In unserem Computer-Programm zur Faktorenanalyse ist der Rippe-Test eingebaut.

Die statistischen Tests führen gelegentlich zu wenig plausiblen Ergebnissen. Die Ursache liegt darin, dass der errechnete Chi-Quadrat-Wert von der Zahl der Untersuchungseinheiten N abhängig ist, N jedoch nicht in die Berechnung der Zahl der Freiheitsgrade eingeht. Bei ein und derselben Korrelationsmatrix, aber verschiedenem N, ändert sich der Chi-Quadrat-Wert proportional zu N. Bei einer Zahl der Untersuchungseinheiten von N = 101 wäre Chi-Quadrat doppelt so groß wie bei N = 51. Die Zahl der Freiheitsgrade bleibt aber unverändert.

P30.1.6.3 Das Kaiser- und das Guttman-Kriterium

Kaiser hat vorgeschlagen, die Faktorenanalyse mit 1.0 in der Diagonalen durchzuführen. Dann sind alle Faktoren mit einem Eigenwert größer 1.0 signifikant. In seiner Alpha-Faktorenanalyse hat Kaiser dieser Regel eine nachträgliche Begründung gegeben.
Guttman (1954) schlägt vor, eine Faktorenanalyse mit den multiplen Bestimmtheitsmaßen R_2 durchzuführen. Er hat nachgewiesen, dass dann die Zahl der Faktoren, die positive Eigenwerte besitzen (die größer .0 sind) die maximale Zahl der signifikanten Faktoren ist. D. h. es kann nicht sein, dass mehr Faktoren tatsächlich signifikant sind. Allerdings können weniger Faktoren signifikant sein.
Das Kaiser-Kriterium ist nicht nur rechnerisch bequem. Es hat sich auch in der Praxis bewährt. Wir müssen allerdings berücksichtigen, dass bei geringer Variablenzahl die Diagonalglieder der Korrelationsmatrix nicht gleich 1.0 gesetzt werden sollen.
Das Guttman-Kriterium gibt in den meisten Fällen zu viele Faktoren als signifikant an. In Prog30m2 wird eine Option angeboten, die Zahl der zu extrahierenden Faktoren zu beschränken auf die Zahl der Eigenwerte größer 1, die aus einer Korrelationsmatrix mit 1.0 in der Diagonalen entstehen. Erst danach werden die multiplen Bestimmtheitsmaße in die Diagonale der Korrelationsmatrix eingesetzt und die Faktorenanalyse gerechnet.

P30.1.6.4 Test auf inhaltliche Interpretierbarkeit

Dieser Test, dem wir den Vorzug geben, geht davon aus, dass nur jene Faktoren akzeptiert werden sollen, die inhaltlich interpretierbar sind.
Die Vorgehensweise ist folgende. Man bestimmt zuerst nach dem Guttman- oder Kaiser-Kriterium die Zahl der signifikanten Faktoren.
Sie sei mit It symbolisiert. Dann wird für It, It–1, It–2, ... 2 Faktoren schiefwinklig rotiert. Die Rotationsergebnisse werden inhaltlich interpretiert. Dabei wird dann eine Entscheidung über die Zahl der inhaltlich signifikanten Faktoren getroffen.

P30.1.7 Die Kommunalitätenschätzung

Je größer eine Matrix ist, um so weniger wichtig wird es, dass man die Kommunalitäten gut schätzt. Bei Matrizen ab ca. 15 Variablen ist die Auswirkung verschiedener Kommunalitätenschätzungen auf die Ergebnisse gering, bei kleineren Matrizen allerdings stark.

Folgende Methoden der Kommunalitätenschätzung sind empfehlenswert:

(1) **In die Hauptdiagonale wird 1.0 gesetzt**.
Die Matrix ist dann prinzipiell positiv semidefinit. Es können alle Faktoren extrahiert werden.
Diese Methode ist nur bei Matrizen von 15 Variablen und mehr empfehlenswert, ist aber für diese wohl die beste Methode, da sie auch die Berechnung der Faktorenwerte, sofern dies vorgesehen ist, begünstigt (siehe Abschnitt P30.1.8) und da sie die Anwendung des einfachen Kaiser-Kriteriums zur Bestimmung der Zahl der signifikanten Faktoren ermöglicht.

(2) **Die Kommunalitäten-Iteration**.
In die Hauptdiagonale wird 1.0 gesetzt. Damit ist die Matrix positiv definit. Es werden die signifikanten Faktoren extrahiert.
Aus den quadrierten Ladungen der Variablen auf den p signifikanten Faktoren ergibt sich ihre Kommunalität.

(1) $$h_i^2 = a_{i1}^2 + \ldots + a_{ij}^2 + \ldots + a_{ip}^2$$

h_i^2 = Kommunalität der Variablen i

a_{ij}^2 = quadrierte Ladung der Variablen i auf Faktor j

Die so berechneten Kommunalitäten werden wieder in die Hauptdiagonale eingesetzt und die Faktorenanalyse wiederholt. Der Nachteil des Verfahrens besteht darin, dass es keinen eindeutigen Test auf Faktorensignifikanz gibt.
Für die Iteration ist es entscheidend, welche Faktorenzahl für die Kommunalitätenberechnung zugrunde gelegt wird.
Man könnte nun bei diesem beschriebenen Verfahren nach der zweiten, wiederholten Faktorenanalyse aus den Ergebnissen neue Kommunalitäten berechnen, diese in die Hauptdiagonale einsetzen, eine dritte Faktorenanalyse durchführen, aus deren Ergebnissen neue Kommunalitäten berechnen, mit diesen eine 4. Faktorenanalyse durchführen usw. ..... dies so lange, bis die eingesetzten Kommunalitäten mit den errechneten übereinstimmen.
Diese Kommunalitäteniteration führt bei kleineren Matrizen erfahrungsgemäß sehr schnell zu Konvergenz. Nicht selten besteht schon etwa nach der 10. Wiederholung nur noch an der 3. Kommastelle eine Differenz zwischen eingesetzter und errechneter Kommunalität.
Dieses Verfahren ist zwar elegant. Es gibt jedoch keinen Beleg dafür, dass die schließlichen Kommunalitäten empirisch "wahrer" sind als die bei der 1. Wiederholung eingesetzten Werte.

Die Kommunalitäteniteration wird auch gelegentlich mit den multiplen Bestimmtheitsmaßen R^2 als Anfangswerte begonnen. Hier gilt unsere Kritik in gleicher Weise. Bei der Kommunalitäteniteration findet nicht notwendigerweise eine Annäherung an die empirisch wahren Werte statt.
Gelegentlich werden durch die Komrnunalitäteniterationen auch Kommunalitäten erzeugt, die größer als 1.0 sind. Damit würd das Modell der Faktorenanalyse verletzt. Almo verwendet deswegen die vorausgegangene Iterationslösung als die endgültige. In Prog30m2.Msk können in der Optionsbox "weitere Optionen" verschiedene Zwischenergebnisse angefordert werden. Dabei werden auch alle Iterationsschritte mitgeteilt. Siehe auch die ausführliche Darstellung in Abschnitt P30.3.6.1.

(3) **Multiple Bestimmtheitsmaße als Diagonalglieder**

Jede einzelne Variable kann mit der zusammengefaßten Menge der m – 1 anderen Variablen korreliert werden. Dabei entsteht der multiple Korrelationskoeffizient R. Das Quadrat dieses Koeffizienten R_i^2 (auch "multiples Bestimmtheitsmaß genannt) ist ein Ausdruck der Gemeinsamkeit, die zwischen der Variablen und den übrigen m – 1 Variablen besteht. Wie Guttman (1956) nachwies, ist R^2 die Untergrenze der Kommunalität. D. h. die Kommunalität kann nicht kleiner sein als R^2. Aufgrund dieser Eigenschaft wurde R^2 gelegentlich als die "empirische Kommunalität" bezeichnet.

Es bestellt weitgehend Einigkeit darüber, dass R^2 die beste Kommunalitätenschätzung ist. Die Berechnung der R^2 erfordert die Inversion der Korrelationsmatrix (mit 1.0 in der Diagonalen).

$$R_i^2 = 1 - \frac{1}{c_{ii}}$$

R_i^2 = multiples Bestimmtheitsmaß der Variablen i

c_{ii} = das i-te Diagonalelement der invertierten Korrelationsmatrix.

Werden die multiplen Bestimmtheitsmaße als Kommunalitätenschätzung eingesetzt, dann kann auch das Guttman-Kriterium zur Bestimmung der Zahl der signifikanten Vaktoren angewendet werden (siehe Abschnitt P30.1.6.3).

P30.1.8 Die Berechnung der Faktorwerte

In P30.1.4, dem Abschnitt über die Rotation, haben wir gezeigt, wie durch die Punktewolken schiefwinkelige Achsen gelegt werden. Diese Achsen, so haben wir gesagt, repräsentieren die latenten Meßdimensionen, die in den Variablen-Untergruppen (die die Punktewolken bilden) „stecken“.
Was wir haben sind die Werte der Untersuchungseinheiten in den einzelnen Variablen der Untergruppen. Was wir suchen sind die Werte der Untersuchungseinheiten in den latenten Meßdimensionen.

Als Schätzgleichung des latenten Wertes einer Untersuchungseinheit v auf der gemeinsamen Dimension $\mathbf{D}_j$, nach Kenntnis seiner empirischen Antworten P auf die m-Variablen einer Gruppe, schreiben wir:

(1) $$D_{jv} = \beta_{j1} \cdot P_{1v} + \beta_{j2} \cdot P_{2v} + \ldots + \beta_{ji} \cdot P_{iv} + \ldots + \beta_{jm} \cdot P_{mv}$$

D_{jv} = "wahrer" Dimensionswert des Untersuchungseinheiten v auf Dimension j,
P_{iv} = Wert der Untersuchungseinheit v auf Variable i. Die P sind standardisiert.
β_{ji} = Gewichtszahl, die die Variable i mit der Dimension j verbindet.

Wir bezeichnen im folgenden die β-Gewichte als "Faktorwertkoeffizienten" oder "Faktor-Betaladungen" und dem Sprachgebrauch der Faktorenanalyse folgend die Dimensionswerte als "Faktorwerte".
Dabei wollen wir aber gleich darauf hinweisen, dass die Faktorwerte aus der rotierten Ladungsmatrix Q berechnet werden.
Wir können Gleichung 1 auch in Matrix-Schreibweise ausdrücken.

(1a) $\mathbf{d'_j} = \boldsymbol{\beta}'_j * \mathbf{P}$

d'_j = Zeilenvektor der Werte der N Untersuchungseinheiten auf der Dimension j

$\beta_j^{'}$ = Zeilenvektor der m Faktor-Betaladungen für die Dimension j
P = m x N – Matrix der (standardisierten) empirischen Werte der m Variablen

Für p Dimensionen wird (l) in Matrix-Schreibweise zu

(2) $$\boldsymbol{D = B \cdot P}$$

D = das ist die p x N Matrix der Faktorwerte.
P = das ist die m x N Matrix der empirischen Werte.
B = das ist die p x m - Matrix der Faktor-Betaladungen.

Die Aufgabe, die uns nun gestellt ist, besteht darin, die Matrix B der Faktor-Betaladungen zu finden.
Die Lösung hängt davon ab, oh wir die Faktorenanalyse mit 1.0 in der Diagonalen oder mit irgendwelchen anderen Kommunalitätenschätzungen durchgeführt haben.

1) Wenn in die Diagonale der Korrelationsmatrix 1.0 eingesetzt wurde, dann unterstellen wir, dass keine spezifischen Dimensionen wirksam sind. Die Faktorwerte D ergeben sich dann gemäß

(3) $$\boldsymbol{D = (K' \cdot K)^{-1} \cdot K' \cdot P}$$

K = das ist die p*m Matrix der achsparallel auf die schiefwinkeligen Achsen projizierten Ladungen.

Der Ausdruck $(K' \cdot K)^{-1} \cdot K'$ ergibt also die für Gleichung 2 notwendige B-Matrix der Faktor-Betaladungen.(=Faktorwert-Koeffizienten)
$K' \cdot K$ ist eine invertierbare symmetrische Matrix.

Für den allerdings unrealistischen Fall, dass die gemeinsamen Dimensionen senkrecht aufeinander stehen (dass **C = I**) ist **K = L**.

Die Matrix $K' \cdot K = L' \cdot L$ ergibt dann die Diagonalmatrix **M** der Eigenwerte. Wenn nur ein Faktor signifikant ist, dann ist allerdings diese Annahme gerechtfertigt. In diesem Falle ist die Faktorwert-Berechnung besonders einfach

(4) $$\boldsymbol{d_1 = \frac{1}{E_1} \cdot 1'_1 \cdot P}$$

wobei E_1= der l . Eigenwert und $1'_1$ = Zeilenvektor des 1. Faktors aus **L** und d_1= Faktorwert der N Untersuchungseinheiten für den 1. Faktor.

Als lineare Gleichung geschrieben ergibt 4

(5) $$d_{1v} = \frac{a_1}{E_1} P_{1v} + \frac{a_2}{E_1} P_{2v} + \ldots \frac{a_1}{E_1} P_{1v} + \ldots \frac{a_{mp}}{E_1} P_{mv}$$

a_i = 1. F'aktorladung (unrotiert) hinsichtlich der Variable i. Wir können festhalten: Unter den beschriebenen Bedingungen sind die Faktor-Betaladungen (=Faktorwert-Keoffizienten) des ersten unrotierten und einzig signifikanten Faktors gleich den durch den Eigenwert dividierten Faktorladungen.
Da wir mit unseren Daten in der Regel beliebige lineare Transformationen durchführen können, (Korrelationskoeffizienten werden dadurch nicht verändert) könnten wir den konstanten Divisor $\frac{1}{E_1}$ auch weglassen. Der Faktorwert des l.

unrotierten Faktors ist dann der gewichtete Gesamtpunktwert aus m Variablen – wobei als Gewichtszahlen die Faktorladungen a, verwendet werden.

2) Wurde in die Diagonale der Korrelationsmatrix nicht 1.0 eingesetzt, dann ist es empfehlenswert, einen der multiplen Regressionsanalyse entsprechenden Ansatz zu verwenden. Aus der multiplen Regressionsanalyse ist folgende Gleichung bekannt:

(6) $\boldsymbol{\beta' = r' \cdot R^{-1}}$

wobei R^{-1} = die Inverse der Korrelationsmatrix der Variablen l, 2... m und r' der Zeilenvektor der Korrelationen der m Variablen mit dem jeweils betrachteten Faktor ist. r' ist also nichts anderes als die jeweilige Spalte aus der Matrix Q (der auf die schiefwinkligen Achsen rechtwinklig projizierten Ladungen). Wir ersetzen also r' durch Q'_j (wobei das Subskript „j" andeuten soll, dass wir nur die j. Zeile diese Matrix benötigen.
Es kann also geschrieben werden

(7) $\boldsymbol{\beta' = Q'_j \cdot R^{-1}}$

β = Vektor der Faktorwert-Koeffizienten (bzw. Faktorbetaladungen)hinsichtlich der Meßdimension j

Wird (7) in (l a) eingesetzt, dann entsteht

(8) $\boldsymbol{d'_j = Q' \cdot R^{-1} \cdot P}$

Das ist die gesuchte Schätzgleichung. Alle Größen der rechten Seite sind bekannt bzw. berechenbar.

P30.1.8.1 Die multiple Faktorbestimmtheit

Überla empfiehlt, nach Berechnung der Faktorwerte die quadrierte multiple Korrelation (multiples Bestimmtheitsmaß) zwischen den Items und dem Faktor zu ermitteln.

Wir wollen diesen Koeffizienten "Faktorbestimmtheit" nennen.

(9) $R_A^2 = \beta_1 \cdot q_1 + \beta_2 \cdot q_2 + ... + \beta_m \cdot q_m$

R_A^2 = Bestimmtheitsmaß für die Dimension A.

$q_1, q_2, \ldots q_m$ = Werte von Q_j, d. h. Korrelationen der Variablen 1,2,3 ,... m mit der Dimension A.

$\beta_1, \beta_2, \ldots \beta_m$ = die entsprechenden Faktor Betaladungen

P30.1.9 Rotation

Wir wollen nochmals einige Gleichungen aus P30.1.4 betrachten. Die Grundgleichung der Faktorenanalyse lautet:

(1) $\boldsymbol{R_h = K \cdot C \cdot K'}$

R_h ist die Matrix der Korrelationen zwischen den Items (mit Kommunalitäten in der Diagonalen).

K ist die Matrix der schiefwinkligen achsparallel projizierten Faktorladungen. (Siehe Zeichnung, P30.1.4.)

C ist die Matrix der Korrelationen zwischen den (schiefwinkligen) Faktoren. Wenn wir die senkrechten Projektionen der Items auf die schiefwinkligen Achsen, also die Q-Matrix suchen, dann lautet die Grundgleichung

(1a) $\boldsymbol{R_h} = \boldsymbol{Q} \cdot (\boldsymbol{C}')^{-1} \cdot \boldsymbol{Q}'$

Die Korrelationsmatrix C kann in eine Transformationsmatrix T aufgelöst werden:

(2) $\boldsymbol{C} = \boldsymbol{T}' \cdot \boldsymbol{T}$

Eingesetzt in Gleichung (1) entsteht dann

(3) $\boldsymbol{R_h} = (\boldsymbol{K} \cdot \boldsymbol{T}') \cdot (\boldsymbol{T} \cdot \boldsymbol{K}')$

Und in (la) eingesetzt entsteht

(3a) $\boldsymbol{R_h} = \boldsymbol{Q} \cdot \boldsymbol{T}^{-1} \cdot (\boldsymbol{T}')^{-1} \cdot \boldsymbol{Q}$

Wir setzen

(4) $\boldsymbol{K} \cdot \boldsymbol{T}' = \boldsymbol{Q} \cdot \boldsymbol{T}^{-1} = \boldsymbol{L}$

(4a) $\boldsymbol{T} \cdot \boldsymbol{K}' = (\boldsymbol{T}')^{-1} \cdot \boldsymbol{Q}' = \boldsymbol{L}'$

L ist die orthogonale Faktorladungsmatrix

In (3) bzw. (3a) eingesetzt entsteht dann

(5) $\boldsymbol{R_h} = \boldsymbol{L} \cdot \boldsymbol{L}'$

Gleichung 5 ist durch das mathematische Verfahren der "Eigenwert-Eigenvektor-Berechnung" lösbar.

Diese Faktorladungsmatrix **L** ist – es sei wiederholt – nur eine von vielen möglichen, die Gleichung 5 befriedigt. Sie liefert jedoch den Bezugsrahmen für das Auffinden der richtigen Matrix.

Die Aufgabe besteht nun darin, von Gleichung (5) zurück zu Gleichung (1) zu gelangen. Das ist die Aufgabe der Rotation. Von "Rotation" wird gesprochen, weil diese Aufgabe dadurch gelöst wird, dass das Koordinatensystem der Matrix **L** in die Position der Matrix **K** rotiert wird.

Der Zweck der Rotation

Die Rotation ist, das geht aus den Gleichungen 1 bis 5 hervor, eine Notwendigkeit. Die Faktorenanalyse liefert eine inhaltlich nicht interpretierbare Matrix **L** von Faktorladungen. Das wird häufig nicht verstanden. (Lediglich für den Sonderfall, dass nur ein Faktor signifikant ist, besteht eine Interpretationsmöglichkeit.) Die Faktorladungsmatrix **L** gibt jedoch ein Bezugssystem ab, aus dem die endgültige, inhaltlich interpretierbare Matrix **K**, bzw. Q relativ leicht gewonnen werden kann – eben durch Rotation.

Der Zweck der Rotation, wenn Variablen-Untergruppen (Punktewolken) analysiert werden sollen, ist also, ausgehend von der "Bezugsmatrix" **L** zu einer Matrix zu rotieren, deren Spalten als eigenständige Meßdimensionen interpretiert werden können. Die Rotation ist demnach nur dann sinnvoll, wenn die Punkte (die die Variablen repräsentieren) mindestens zwei signifikant voneinander getrennte und in sich relativ geschlossene Punktwolken bilden.

Unser Rotationsprogramm
Unsere Vorgehensweise bei der Lösung des Rotationsproblems ist folgende

1) Zuerst gilt es, die Punktwolken (d. h., die Gruppen der zusammengehörenden Variablen) zu isolieren. Wir nennen diesen Schritt "Gruppenbildung". Wir verwenden dabei das Quartimin-Verfahren. Es ist auch ein eigenständiges Rotationsverfahren. Es ist nicht nur für die Gruppenbildung verwenbar. Dies gilt vor allem dann, wenn keine voneinander trennbaren Punktewolken identifizierbar sind.
2) Durch die identifizierten Punktwolken werden Achsen gelegt – und zwar in der Weise, dass sie ein Maximum an Varianz je Punktwolke erklären. Damit entsteht die gesuchte Matrix **K** bzw. **Q**. Die Methode, die vom Verfasser für diese Aufgabe eigens entwickelt wurde, wird "Gruppenrotation" genannt.

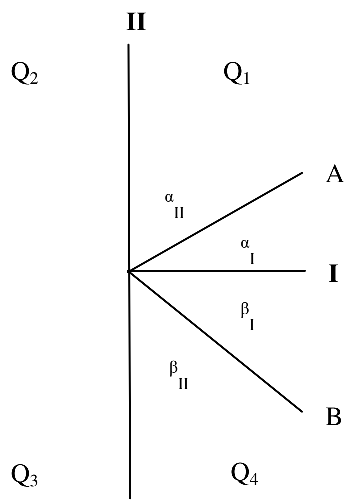
P30.1.9.1 Die Transformationsmatrix

Die Aufgabe der Rotation – so wurde gesagt – besteht darin, von der Matrix **L** zur Matrix **K** oder **Q** zu gelangen.
Dazu ist es notwendig, die Transformationsmatrix **T** zu finden.

Durch Umordnung von Gleichung (4) erhalten wir

(6) $\boldsymbol{K} = \boldsymbol{L} \cdot (\boldsymbol{T}')^{-1}$
(7) $\boldsymbol{Q} = \boldsymbol{L} \cdot \boldsymbol{T}$

Wenn wir **L** mit der Transformationsmatrix **T** postmultiplizieren, erhalten wir **Q**, die Matrix der rechtwinkligen Projektionen der Items auf die schiefwinkligen Achsen. Wird **L** mit $\mathbf{(T')^{-1}}$ postmultipliziert, dann entsteht K, die Matrix der achsparallelen Projektionen. Wie sieht nun die Transformationsmatrix aus? Wir wollen das am Beispiel einer Variablenmenge verdeutlichen, die aus 2 Punktewolken (mit den Achsen A und B) besteht.



Wenn der Winkel zwischen den beiden neuen Achsen A und B 90° beträgt, dann wird von orthogonaler Rotation gesprochen. Die Achsen I – II werden in die Position A – B rotiert. Wird die Bedingung, dass zwischen A und B ein rechter Winkel bestehen muss fallengelassen, dann wird von schiefwinkliger Rotation gesprochen.
Transformationsmatrix

		neue Achsen A	B
"alte orthogonale	I	cos αI	cos βI
Achsen"	II	cos αII	cos βII

Die Transformationsmatrix enthält die Kosinuswerte der Winkel zwischen der neuen Achse und den beiden alten orthogonalen Achsen. Dabei ist zu beachten, dass das Vorzeichen der Kosinuswerte positiv ist, wenn die neuen Achsen im Quadranten Q 1 und Q 4 liegen, jedoch negativ, wenn sie in Q 2 und Q 3 liegen.
Besitzt die **L**-Matrix drei signifikante Faktoren, dann ist die **T**-Matrix:

		neue Achsen A	B	C
alte orthogonale	I	cos αI	cos βI	cos γI
Achsen	II	cos αII	cos βII	cos γII
	III	cos αIII	cos βIII	cos γIII

Im 3-dimensionalen Fall kann man sich die alten Achsen I, II und III als ein Dreibein vorstellen. In diesem Raum, aus demselben Ursprung hervorgehend, verläuft die Achse A. Sie besitzt drei "durch den Raum gehende" Winkel gegenüber den 3 alten Achsen. Der Kosinus dieses Winkels wird "Richtungs-Kosinus" genannt.
In der Transformationsmatrix sind die Richtungs-Kosinuswerte dieser 3 Winkel mit $\cos\alpha_{I}, \cos\alpha_{II}$ und $\cos\alpha_{III}$ enthalten. Entsprechendes gilt für die anderen neuen Achsen B und C.

Zwei wichtige Sätze sind festzuhalten:

1. Die Transformationsmatrix ist eine quadratische p x p Matrix. Wenn die L-Matrix p Spalten enthält, dann muss auch die K- bzw. Q-Matrix p Spalten enthalten.
2. Die quadrierten Richtungskosinuswerte einer neuen Achse A gegenüber allen alten Achsen addieren sich auf zu 1.0.

Was wir hier für den 2- und 3-dimensionalen Fall vorgetragen haben, gilt allgemein für den p-dimensionalen Fall. **T** ist dann eine p x p Matrix, die in ihren Spalten die Richtungskosinuswerte der neuen Achsen gegenüber den alten orthogonalen enthält.

Transformationsmatrix bei orthogonaler Rotation:
Wird bei der Drehbewegung zwischen den neuen Achsen A und B ein rechter Winkel festgehalten (orthogonale Rotation), dann vereinfacht sich die Transformationsmatrix erheblich.

Es gilt dann:

(8) $\cos \alpha_{II} = 90° - \alpha_{I}$
(9) $\beta_{II} = \alpha_{I}$
(10) $\beta_{I} = 90° - \alpha_{I}$

Daraus folgt für die Kosinuswerte der Transformationsmatrix:

(8a) $\cos \alpha_{II}=\cos(90-\alpha_I)=\sin \alpha_I$
(9a) $\cos \beta_{II}=\cos \alpha_I$
(10a) $\cos \beta_I = \cos(90-\alpha_I)=\sin \alpha_I$
Wir müssen noch die Vorzeichen der Kosinus- und Sinuswerte einsetzen. Dabei muss unterschieden werden, ob das neue orthogonale System A – B entgegen oder mit dem Uhrzeigersinn im alten System I – II um den Winkel α_I gedreht wird.

entgegen Uhrzeigersinn

$\cos \alpha$	$-\sin \alpha$
$\sin \alpha$	$\cos \alpha$

mit Uhrzeigersinn

$\cos \alpha$	$\sin \alpha$
$-\sin \alpha$	$\cos \alpha$

(Der Einfachheit halber haben wir das Subskript I vom Winkel α weggelassen.)

P30.2 Eingabe und Ausgabe in einfaches Programm 30

P30.2.1 Eingabe in Programm-Maske Prog 30m1

Prog30m1 erfordert nur eine minimale Eingabe. Sie finden dieses Programm, wenn Sie auf den Knopf „Verfahren“ klicken und in der dann präsentierten Auswahlliste „Faktorenanalyse“ selektieren.

Voreinstellungen im Maskenprogramm Prog30m1

Im Maskenprogramm sind folgende Voreinstellungen gültig.

1. In einem ersten Schritt wird aus Daten die Korrelationsmatrix als Ausgangsmatrix für die Faktorenanalyse errechnet.
2. Als Kommunalitäten wird 1.0 in die Diagonale eingesetzt.
3. Es werden alle Faktoren extrahiert, die einen Eigenwert größer 1.0 besitzen.
4. Für diese Faktoren wird eine schiefwinkelige Rotation durchgeführt.

Prog30m1.Msk
Faktorenanalyse

Beispiel: Es soll festgestellt werden, ob 6 verschiedene sportliche Leistungen auf einer bzw. mehreren Fähigkeiten (=Faktoren) beruhen.

Aus den Variablenwerten der Untersuchungspersonen wird zuerst eine Korrelationsmatrix gebildet. Diese wird dann mit der Hauptkomponenten-Methode (d.h. mit 1.0 in der Diagonalen) faktorisiert. Es entsteht die unrotierte Faktorladungsmatrix. Diese kann als 2- bzw. 3-dimensionales Koordinatensystem grafisch dargestellt werden. Dann wird schiefwinklig rotiert. D.h. es wird versucht, räumlich voneinander getrennte "Punktewolken" zu identifizieren und durch diese Achsen zu ziehen. Diese Achsen können dann als die geometrischen Repräsentanten der (in unserem Beispiel) Fähigkeiten betrachtet werden, die hinter den 6 sportlichen Leistungen stehen. In unserem Beispiel könnten diese mit "Kraft", "Schnelligkeit" und "Ausdauer" benannt werden

Siehe Handbuch, Abschnitt P30

Programm-Bedienung ---> Hilfe

Speicher fuer x Variable Hilfe

Vereinbare Variable= 20

Option: Weitere Vereinbarungen - nur wenn Almo dazu auffordert

Variablennamen

Datei der Variablennamen

Hilfe

zeige = Namensdatei in Output zeigen
leer = nicht zeigen

Freie Namensfelder

Hilfe

Leere alle Eingabefelder dieser Sub-Box

Name1=Marathon
Name2=Rad
Name3=Kugelstoss
Name4=Speerwurf
Name5=Weitsprung
Name6=Kurzstrecke

erzeuge zusätzliche Namensfelder

Variablennamen in Datei speichern
Eingabefeld leer = nicht speichern

Hilfe

Datei aus der gelesen wird

Hilfe
bei Datei-Problemen

".\Testdat\Fakdat.fre"

frei

Format der Daten

Hilfe

V1:6

der Datensatz enthält diese Variablen
Bei Format DIREKT schreiben Sie: alle_V

zu faktorisierende quantitative Variable

Marathon,Rad,Kugelstoss,Speerwurf

Option: Umkodierungen und Kein-Wert-Angaben

Grafik-Optionen

Programmende

Erläuterungen zu den Dialogboxen

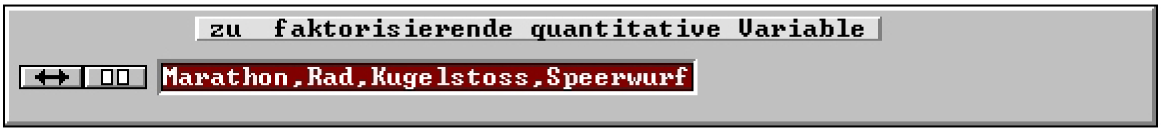
Eingabebox: Speicher für x Variable
Siehe dazu Almo-Dokument Nr. 0 "Arbeiten mit Almo", Abschnitt P0.1.
Die Zahl der vereinbarten Variablen muss mindestens so hoch sein, wie die höchste im gesamten Programmtext vorkommende Variablennummer.
Normalerweise ist dies die Nummer der letzten Variable des eingelesenen Datensatzes.
Sie können die Zahl der vereinbarten Variablen aus Sicherheitsgründen auch höher setzen.
Beispiel: Ein Datensatz aus Ihrer Datei umfasst 40 Variable. Dann geben Sie mindestens 40 an. Siehe Teil 2, Abschnitt 35

Eingabebox: Weitere Vereinbarungen
Siehe dazu dazu Almo-Dokument Nr. 0 "Arbeiten mit Almo", Abschnitt P0.2.

Eingabebox: Datei der Variablennamen
Eingabebox: Freie Namensfelder
Zu diesen beiden Eingabeboxen siehe P0.3.

Eingabebox: Datei aus der gelesen wird
Eingabebox: Wenn Dateiformat FIX oder Nicht-Standard-FREI
Zu diesen beiden Eingabeboxen siehe P0.4.

Eingabebox: Die zu faktorisierenden Variablen



Siehe dazu Almo-Dokument Nr. 0 "Arbeiten mit Almo", Abschnitt P0.11.
Klicken Sie zu auf den Knopf mit den zwei kleinen Fenstersymbolen. Almo präsentiert Ihnen dann die Variablen-Auswahl-Box. In ihr klicken Sie auf die Variablen, die faktorisiert werden sollen.

Eingabebox: Umkodierung und Kein_Wert-Angabe und Umkodierungen
Siehe P0.5.

P30.2.3 Ausgabe

P30.2.3.1 Beispiel mit 2 Faktoren

ALMO gibt beim Maskenprogramm und beim äquivalenten selbstgeschriebenem Programm zunächst (wie bei P19 im Teil 3) die Korrelationsmatrix aus und dann die nachfolgenden Ergebnisse der Faktorenanalyse und der Rotation.

Korrelations-Matrix

		Marathon V1	Rad V2	Kugelsto V3	Speerwurf V4
Marathon	V1	1.0000	0.3974	0.5281	0.2534
Rad	V2	0.3974	1.0000	0.0945	0.0305
Kugelstoss	V3	0.5281	0.0945	1.0000	0.7442
Speerwurf	V4	0.2534	0.0305	0.7442	1.0000

Mindestgroesse des Produkt-Moment- Korrelationskoeffizienten r	bei Signifikanz (1-p)*100	fuer df=n-2=50-2=48
0.1844	80	
0.2067	85	
0.2355	90	
0.2541	92.5	
0.2786	95	
0.3168	97.5	
0.3603	99	
0.4542	99.9	

==

Ergebnisse aus Faktorenanalyse
----

2 Eigenwerte der Korrelationsmatrix sind groesser 1.0
Entsprechend versucht Almo 2 Faktoren fuer die nachfolgende Faktorenanalyse zu extrahieren

Eingesetzte Werte fuer Kommunalitaeten

Marathon	V1	1.0000
Rad	V2	1.0000
Kugelstoss	V3	1.0000
Speerwurf	V4	1.0000

********** ***Erläuterung*:**
Almo rechnet standardmäßig mit Kommunalitäten von 1.0. Dann werden nur jene Faktoren betrachtet, die einen Eigenwert von größer 1.0 haben. Das ist das sogenannte Kaiser-Kriterium. Diese Voreinstellung kann vom Benutzer verändert werden. Siehe dazu das nächste Maskenprogramm Prog30m2.Msk

Zahl der Kommunalitaeten-Iterationen: 0

Koeffizienten fuer Faktoren

Eigenwerte (Varianz je Faktor)
2.1125 1.1553

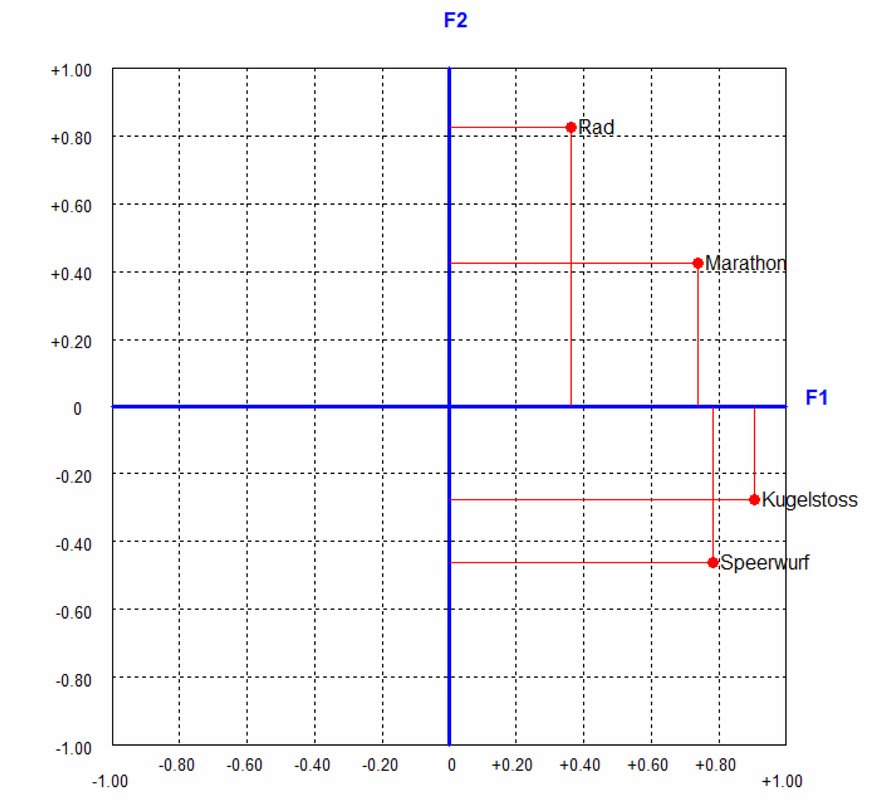
Prozent der Varianz je Faktor
52.8114 28.8816

Zu erklaerende Gesamtvarianz= 4.0000
Durch 2 Faktoren erklaerte Varianz= 3.2677
Prozentsatz der erklaerten Varianz= 81.6929

Matrix der Faktorladungen

		Faktor 1	Faktor 2
Marathon	V1	0.7383	0.4243
Rad	V2	0.3623	0.8272
Kugelstoss	V3	0.9060	-0.2752
Speerwurf	V4	0.7843	-0.4637

Die Faktorladungsmatrix wird von Almo grafisch dargestellt




```
Kommunalitaeten je Variable
```

Marathon	V1	0.7251
Rad	V2	0.8157
Kugelsto	V3	0.8966
Speerwur	V4	0.8301

********** ***Erläuterung***:
Die tatsächlichen Kommunalitäten werden ermittelt. Sie ergeben sich als Summe der quadrierten Faktorladungen

P30.2.3.2 Schiefwinklige Rotation mit 2 Faktoren

Almo liefert folgende Ausgabe:

```
Quartimin-Kriterium                                    0.1776
letzte Iterationsdifferenz bei Quartimin-Rotation      0.0000

Variablengruppe 1: V3 V4
Variablengruppe 2: V1 V2

Durch zugehoerige Achse erklaerte Varianz je Variablengruppe
  1.7023    1.3907
```

********** ***Erläuterung***:
In 3 aufeinder folgenden Schritten geschieht folgendes:

1. Schritt: Almo rotiert schiefwinklig wobei das Quartim-Verfahren verwendet wird. In unserem Beispiel wurden 2 Faktoren extrahiert.

2. Schritt: Danach wird versucht , die Variablen, entsprechend ihrer räumlichen Nähe zu den 2 Faktoren, in 2 Variablengruppen einzuteilen.

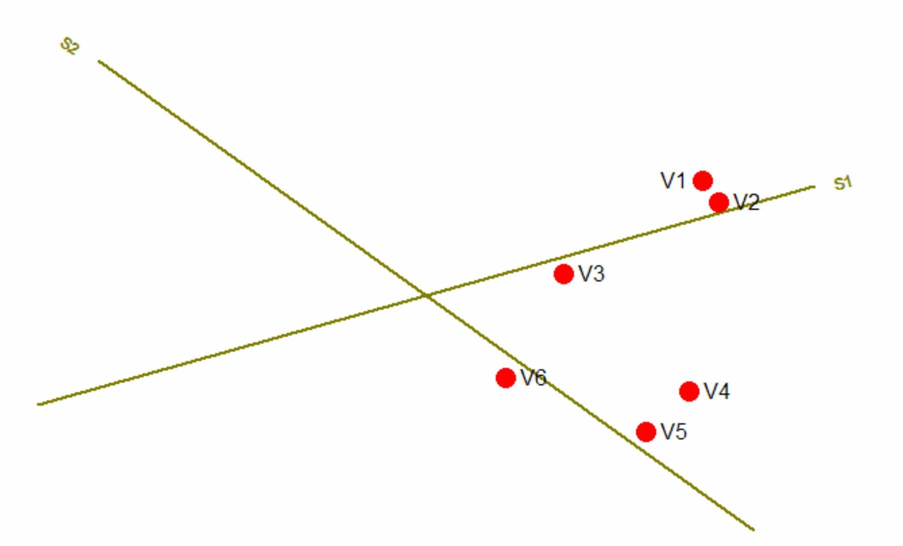
3. Schritt: Danach wird die Quartimin-Lösung durch das "Gruppenrotations-Verfahren" verbessert, indem varianzmaximierende Achsen durch die beiden Variablengruppen hindurch gelegt werden.

Man erkennt, in obiger Grafik, zwei "Punktewolken. Die eine Punktewolke besteht aus Rad und Marathon, die andere aus Kugelstoßen und Speerwurf. Wir erkennen schon, wie die beiden Punktwolken zu benennen sein werden.Die erste ist "Ausdauer", die zweite ist "Kraft"

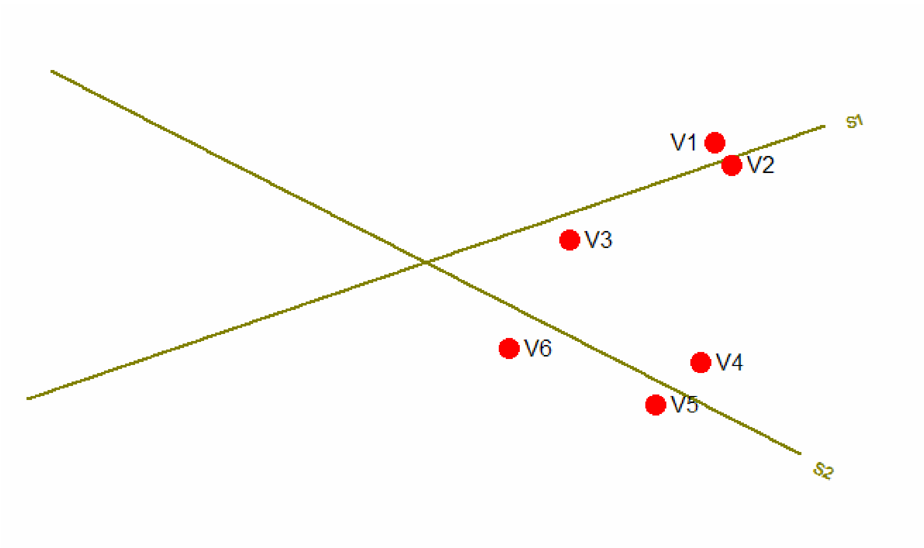
Das **Gruppenrotationsverfahren** verläuft in folgenden Schritten:

1. Aus den Ergebnissen der Quartimin-Rotation werden soviele "Variablengruppen gebildet wie Faktoren extrahiert wurden. Eine Variable wird der Gruppe zugeordnet, zu deren (schiefwinklige) Quartimin-Achse sie den geringsten Abstand besitzt. Ein Variablengruppe ist also grafisch betrachtet eine Punktewolke um eine schiefwinklige Achse herum.

2. Das Gruppenrotation-Verfahren besteht nun einfach darin, dass die Quartimin-Achse etwas gedreht wird. Sie wird so gedreht dass die Summe der quadrierten Ladungen der zu ihr gehörenden Variablen ein Maximum wird. In der Regel muss dabei die Quartimin-Achse nur um einen kleinen Winkel gedreht werden. Demzufolge sind auch die Ergebnisse aus der Quartimin- und der Gruppen-Rotation sehr ähnlich.
Beispiel: Die Quartimin-Rotation hat - grafisch dargestellt - folgendes Ergebnis erbracht:



Die Punkte V1,2,3 bilden eine Punktewolke um die schiefwinklige Achse S1. Die Punkte V4,5,6 bilden eine Punktewolke um die schiefwinklige Achse S2. Bei der Quartimin-Rotation wird - sehr vereinfacht ausgedrückt - versucht die Achsen so zu legen, dass eine "Einfachststruktur" entsteht.
Die nachfolgende Grafik zeigt das Ergebnis nach der Gruppenrotaion. Es ist gut zu erkennen, dass die Achse S2 nun zwischen den Punkten V4 ubd V5 verläuft - also "mitten" durch die Punktwolke. Die beiden Variablen V4 und V5 haben größere Ladungen und bestimmen dadurch die Lage von S2 stärker. Entsprechendes gilt für V1 und V2 und die Achse S1.



In Prog30m2 mit Optionen (Abschnitt P30.3.8) kann der Benutzer entscheiden, ob er
1. mit dem Quartimin-Verfahren schiefwinklig rotieren will oder ob er
2. das Quartimin-Verfahren noch durch das Gruppenrotations-Verfahren verbessern will
3. oder ob er zusätzlich zum verbesserten Quartimin-Verfahren noch das alleinige unverbesserte Quartimin-Verfahren ausgeben will.

Die Verbesserung des Quartimin-Verfahrens durch das Gruppenrotationsverfahren setzt voraus, dass Vatiablengruppen, d.h. voneinander getrennte Punktewolken identifizierbar sind. Ist das der Fall, dann liefert dieses Verfahren eine Lösung, bei der ein Maximum an Varianz je Variablengruppe erklärt wird. Dies ist ein sehr sinnvolles Kriterium. Ist dies jedoch nicht der Fall, dann liefert dieses Verfahren eine Lösung, die geringfügig anders ist als die reine Quartimin-Lösung, aber nicht richtiger oder falscher als diese ist. Die schiefwinkligen Achsen sind gegenüber denen des Quartimin-Verfahrens etwas verschoben.

```
Matrix der Korrelationen zwischen
den schiefwinkligen Achsen
```

	Faktor 1	Faktor 2
Faktor 1	1.0000	0.2951
Faktor 2	0.2951	1.0000

```
Matrix der Winkel zwischen den schiefwinkligen Achsen
```

	Faktor 1	Faktor 2
Faktor 1	0	72.8367
Faktor 2	72.8367	0

********** ***Erläuterung***:
Durch die beiden Punktewolken werden Achsen gelegt. Die beiden Achsen stehen in einem Winkel von 72.8 Grad aufeinander. Siehe nachfolgende Grafik. Der Kosinus dieses Winkels kann als die Korrelation (von 0.295) zwischen diesen beiden Achsen, d.h. den sportlichen Hintergrundsfaktoren "Ausdauer" und "Kraft" betrachtet werden.

```
Transformstionsmatrix
von orthogonaler Faktorladungsmatrix zu nachfolgender Strukturmatrix
```

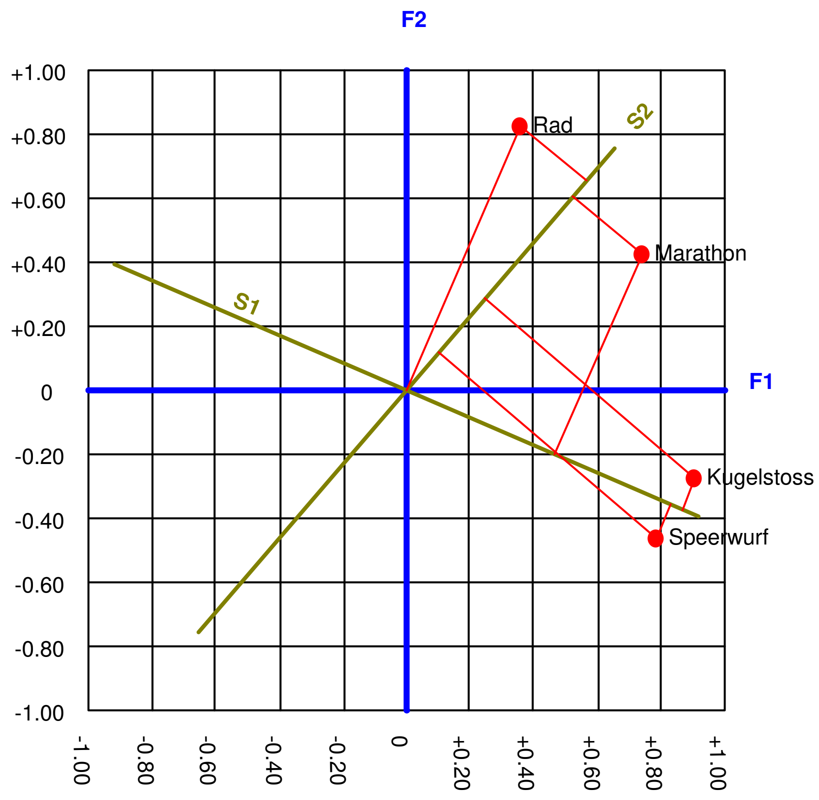
	Faktor 1	Faktor 2
Faktor 1	0.9172	0.6513
Faktor 2	-0.3983	0.7588

Matrix der auf die schiefwinkligen Achsen rechtwinklig projizierten Faktorladungen (Strukturmatrix)

		Faktor 1	Faktor 2
Kugelsto	V3	0.9406	0.3812
Speerwur	V4	0.9041	0.1589
Marathon	V1	0.5081	0.8028
Rad	V2	0.0028	0.8637

Almo erzeugt folgende Grafik für diese Matrix

Faktorladungen im recht- und schiefwinkligen Koordinatensystem (rechtwinklige Projektion)



********** ***Erläuterung***:
Bei dieser Grafik werden die Punkte rechtwinklige auf die schiefwinkligen Achsen projiziert. Dies ist nicht besonders anschaulich

Transformstionsmatrix
von orthogonaler Faktorladungsmatrix zu nachfolgender Ladungsmatrix

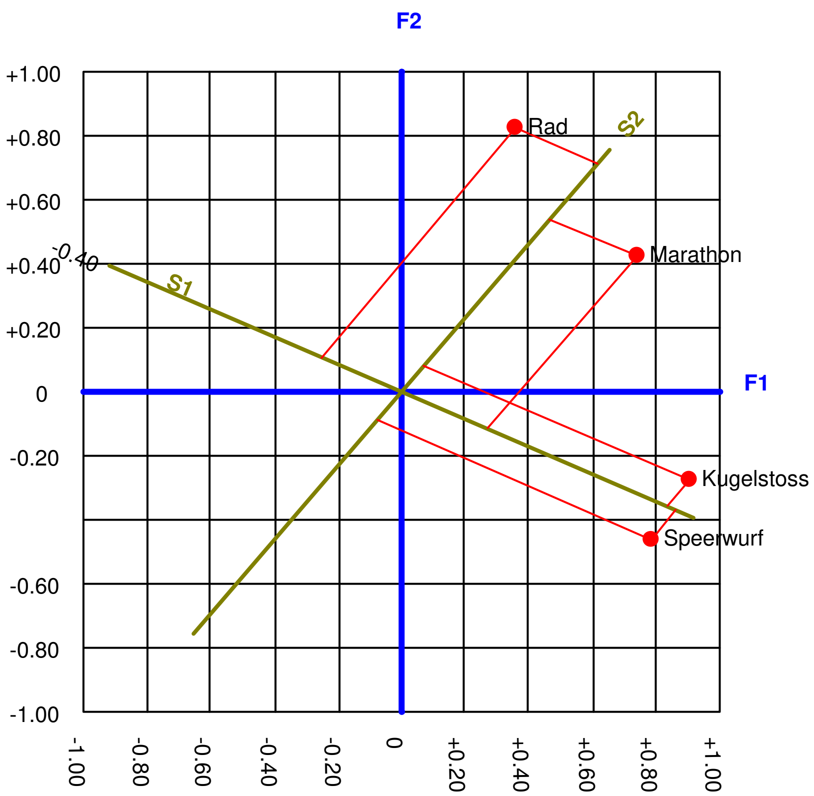
	Faktor 1	Faktor 2
Faktor 1	0.7941	0.4169
Faktor 2	-0.6816	0.9599

Matrix der auf die schiefwinkligen Achsen
achsparallel projizierten Faktorladungen
(Ladungsmatrix)

		Faktor 1	Faktor 2
Kugelsto	V3	0.9071	0.1135
Speerwur	V4	0.9389	-0.1181
Marathon	V1	0.2970	0.7152
Rad	V2	-0.2761	0.9452

Almo erzeugt folgende Grafik für diese Matrix

Faktorladungen im recht- und
schiefwinkligen Koordinatensystem
(achsparallele Projektion)



********** ***Erläuterung*:**

Bei dieser Grafik werden die Punkte achsparallel auf die schiefwinkligen Achsen projiziert. Dies ist die anschaulichste Darstellungsweise. Es ist nun gut zu erkennen, dass zwei "Punktewolken" voneinander getrennt werden können. Die eine Punktewolke besteht aus Rad und Marathon, die andere aus Kugelstoßen und Speerwurf. Wir können die beiden Achsen als geometrische Repräsentanten der beiden sportlichen Hintergrundsfaktoren "Ausdauer" und "Kraft" betrachten.

```
Matrix der raeumlichen Entfernungen
der Variablen von den schiefwinkligen Achsen
```

		Faktor 1	Faktor 2
Kugelsto	V3	0.1084	0.8667
Speerwur	V4	0.1128	0.8971
Marathon	V1	0.6833	0.2838
Rad	V2	0.9031	0.2638

********** ***Erläuterung*****:**
Die (rechtwinklige) Entfernung der Punkte von den Achsen wird ausgegeben.

```
Raeumliche Geschlossenheit der Variablengruppen
(mittlere Entfernung der Variablen einer Variablengruppe
 von "ihrer" Achse)

  0.1106    0.2738
```

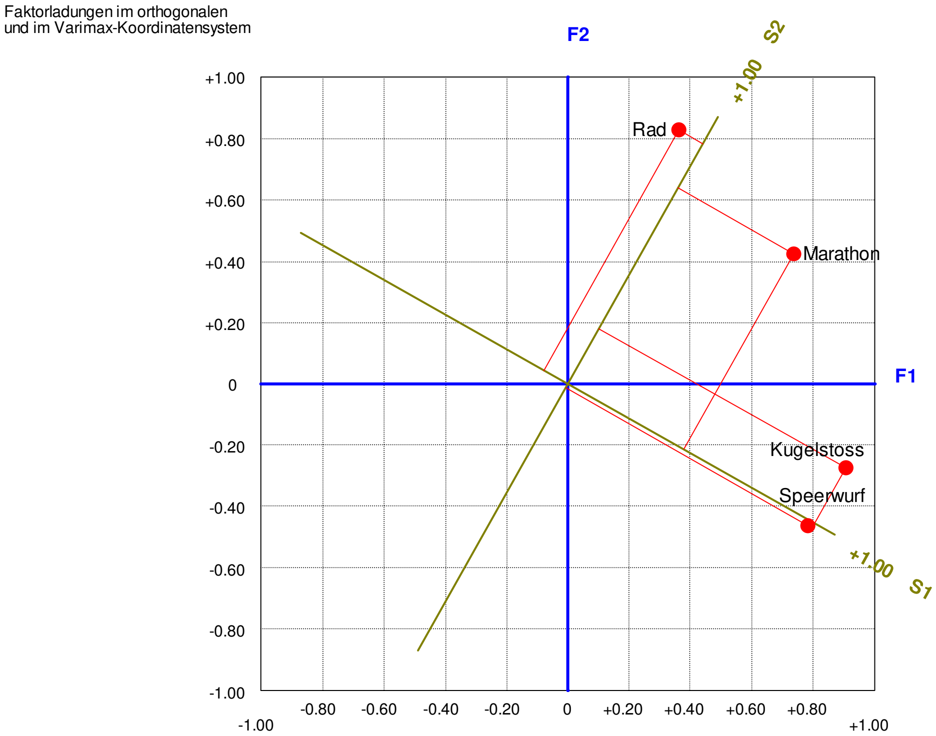
P30.2.3.3 Rechtwinklige Rotation mit 2 Faktoren

Wir hätten anstelle der schiefwinkeligen Rotation auch rechtwinklig rotieren können. Mit dem nachfolgenden Programm Prog30m2 (nicht aber mit Prog30m1) wird dies möglich sein. Wir wollen aber schon hier die Ausgabe dieser Rotationsart zeigen und erläutern. Siehe dazu auch Abschnitt P30.3.6. Die Ergebnisse sind folgende:

```
Rechtwinklig Varimax-rotierte Faktorladungen
```

		Faktor 1	Faktor 2
Marath	V1	0.4345	0.7323
Rad	V2	-0.0907	0.8985
Kugels	V3	0.9243	0.2053
Speerw	V4	0.9109	-0.0185

ALMO liefert folgende Grafik:

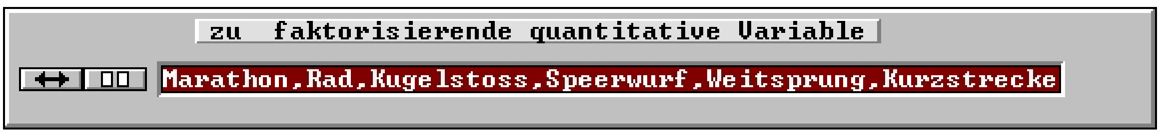


Das ursprüngliche Koordinatensystem ist mit F1-F2 bezeichnet, die rechtwinkligen Achsen mit S1-S2.
Es ist deutlich zu erkennen, dass S1 und S2 keine gute Repräsentanten der beiden Faktoren "Kraft" und "Ausdauer" sind. Die Achsen laufen nicht "mitten durch" die beiden Punktewolken. Die schiefwinklige Rotation ist vorzuziehen.

Almo bietet nun auch noch eine "**benutzerdefinierte rechtwinklige Rotation**" an. Der Benutzer kann dabei die Punktekonfiguration im Koordinatensystem beliebig drehen. Wir werden das in Abschnitt P30.2.3.7 ausführlich beschreiben.

P30.2.3.4 Beispiel mit 3 Faktoren:

Beim Maskenprogramm Prog30m1.Msk müßte nun in der Box "zu faktorisierende Variable" "Weitsprung" und "Kurzstrecke" hinzugefügt werden.



Dadurch entsteht dann eine 3-faktorielle Lösung.

Wir betrachten nur die wesentlichen Teile aus der Almo-Ausgabe

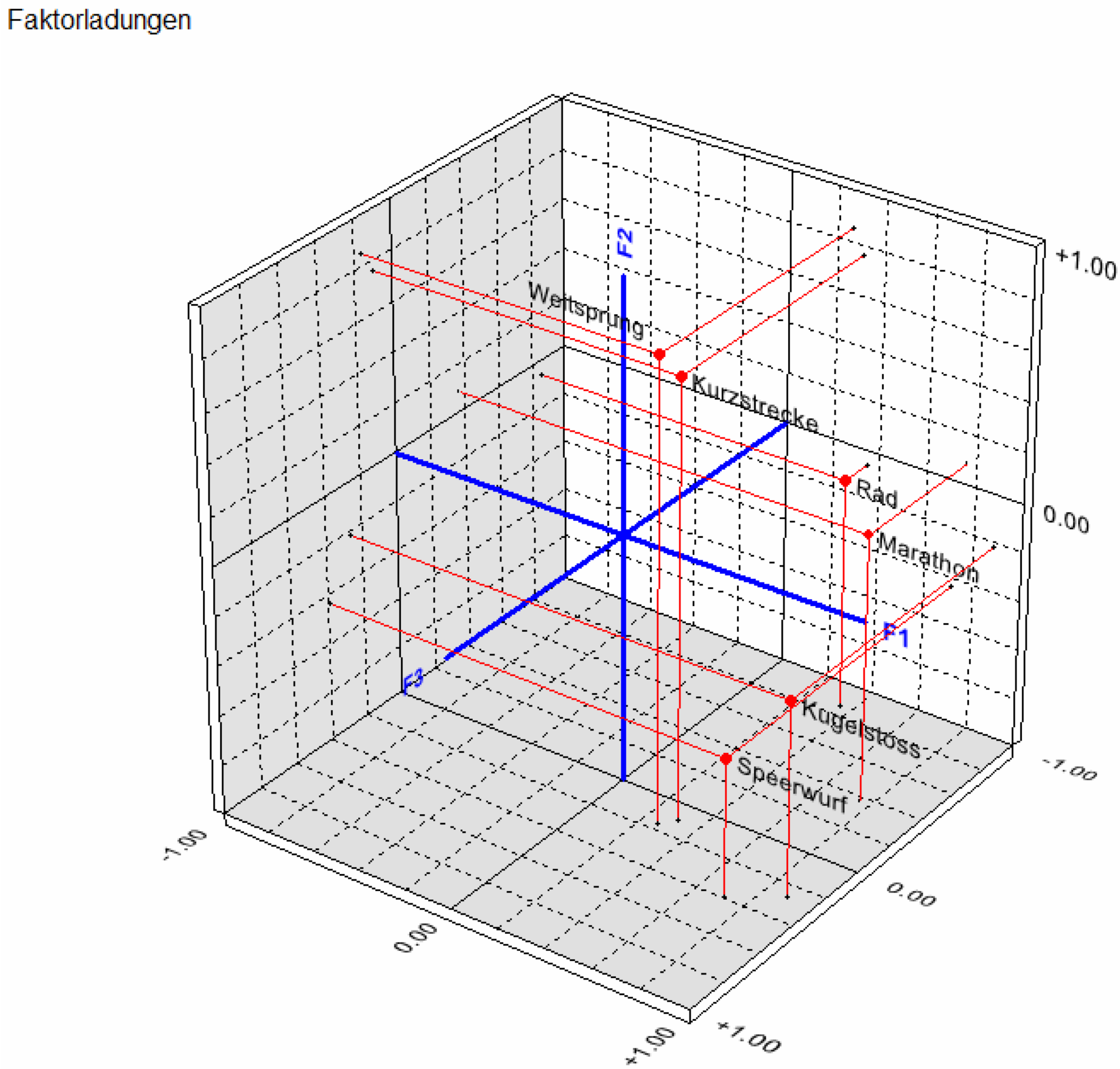
```
Ergebnisse aus Faktorenanalyse
------------------------------
3 Eigenwerte der Korrelationsmatrix sind groesser 1.0
Entsprechend versucht Almo 3 Faktoren fuer die nachfolgende
Faktorenanalyse zu extrahieren

Koeffizienten fuer Faktoren

Eigenwerte (Varianz je Faktor)
    2.1629      1.5529      1.1419
```

```
Prozent der Varianz je Faktor
   36.0483     25.8825     19.0318

Zu erklaerende Gesamtvarianz=         6.0000
Durch 3 Faktoren erklaerte Varianz=  4.8578
Prozentsatz der erklaerten Varianz= 80.9625
```

Matrix der Faktorladungen

		Faktor 1	Faktor 2	Faktor 3
Marathon	V1	0.7611	0.0836	-0.3750
Rad	V2	0.3479	-0.0615	-0.8624
Kugelsto	V3	0.8855	-0.2152	0.2636
Speerwur	V4	0.7153	-0.4377	0.3924
Weitspru	V5	0.2641	0.8526	0.1575
Kurzstre	V6	0.3110	0.7597	0.0954

Almo liefert folgende Grafik



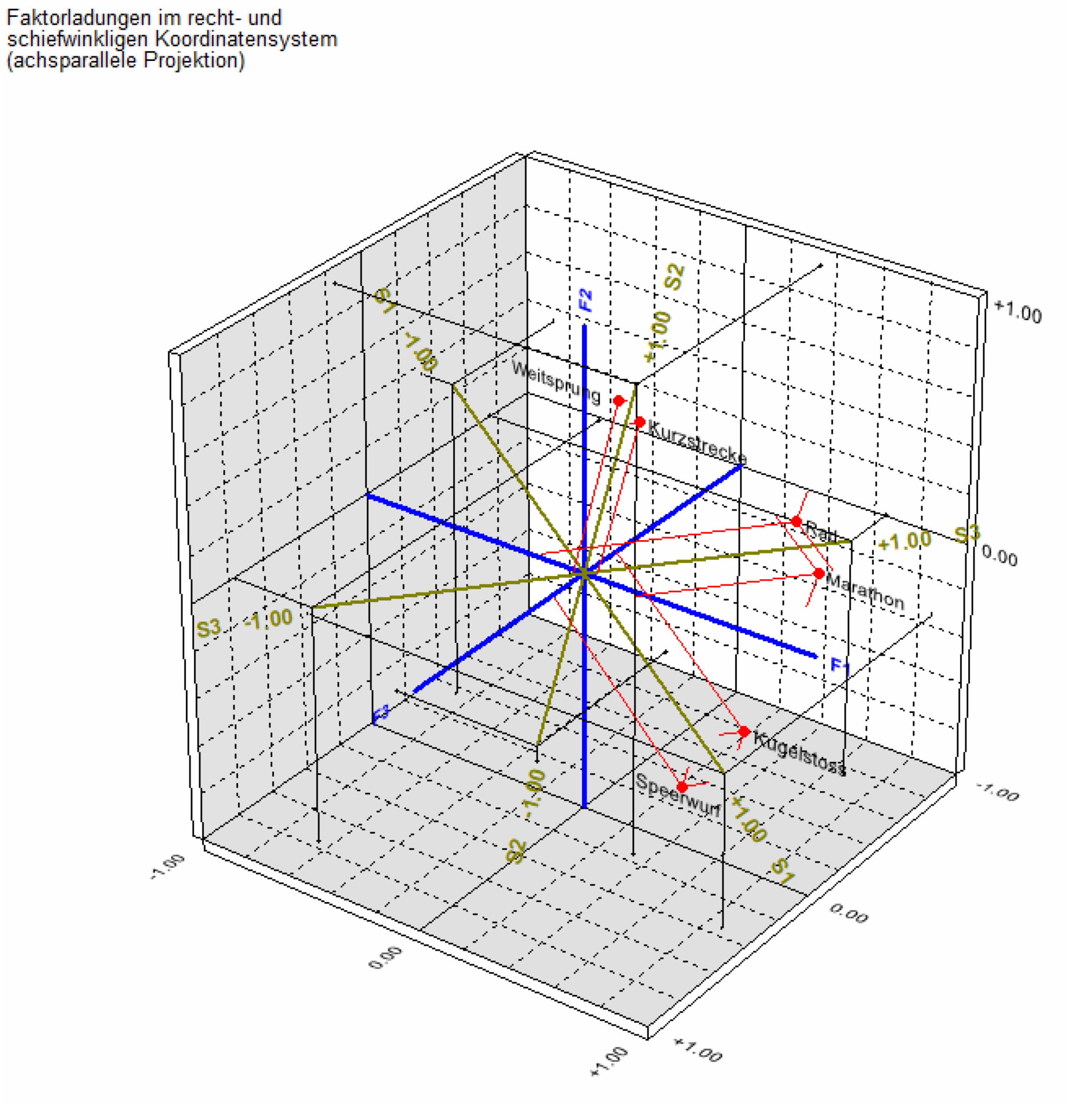
Es ist zu erkennen, dass 3 Punktewolken deutlich gegeneinander getrennt sind. Wir können sie bezeichnen, als "Ausdauer", "Kraft" und "Schnelligkeit".

P30.2.3.5 Schiefwinklige Rotation mit 3 Faktoren

```
Variablengruppe 1: V3 V4
Variablengruppe 2: V5 V6
Variablengruppe 3: V1 V2

Matrix der auf die schiefwinkligen Achsen
achsparallel projizierten Faktorladungen
(Ladungsmatrix)
```

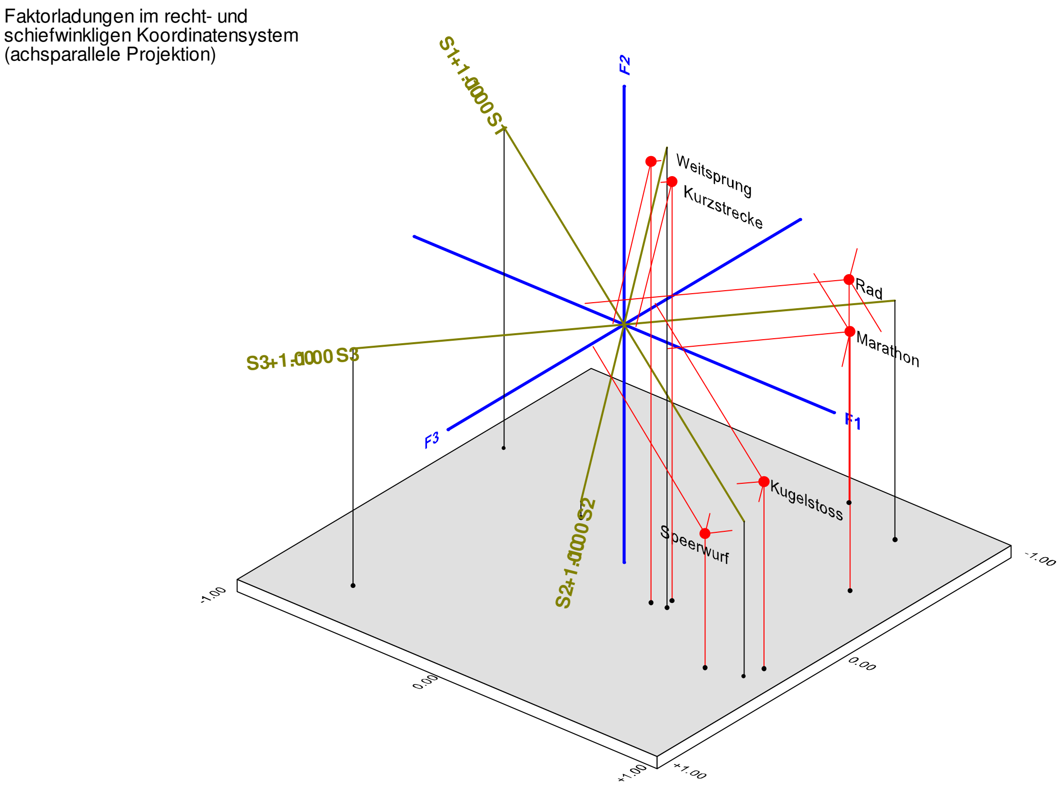
		Faktor 1	Faktor 2	Faktor 3
Kugelsto	V3	0.9070	0.1103	0.0968
Speerwur	V4	0.9430	-0.1131	-0.0993
Weitspru	V5	-0.0141	0.9096	-0.0372
Kurzstre	V6	0.0154	0.8204	0.0408
Marathon	V1	0.2994	0.1969	0.6713
Rad	V2	-0.2653	-0.1745	0.9776

```
Almo liefert folgende Grafik
```



Die Grafik erfordert vom Benutzer ein gutes räumliches Vorstellungsvermögen. In der Grafik sind sowohl die ursprünglichen orthogonalen Achsen F1, F2 und F3 aus der nicht rotierten Lösung enthalten als auch die schiefwinkligen Achsen S1, S2 und S3. Die 6 Punkte sind auf die Flächen projiziert, die diese schiefwinkligen Achsen bilden. Es ist zu erkennen, dass 3 Punktewolken deutlich gegeneinander getrennt sind. Wir können sie bezeichnen, als "Ausdauer", "Kraft" und "Schnelligkeit".

Durch Klick im Output auf den großen Knopf "Grafik" gelangt man in den Almo-Grafik-Editor. Dort kann man versuchen, die Grafik so erscheinen zu lassen, dass man die räumliche Struktur besser erkennt - vielleicht so:



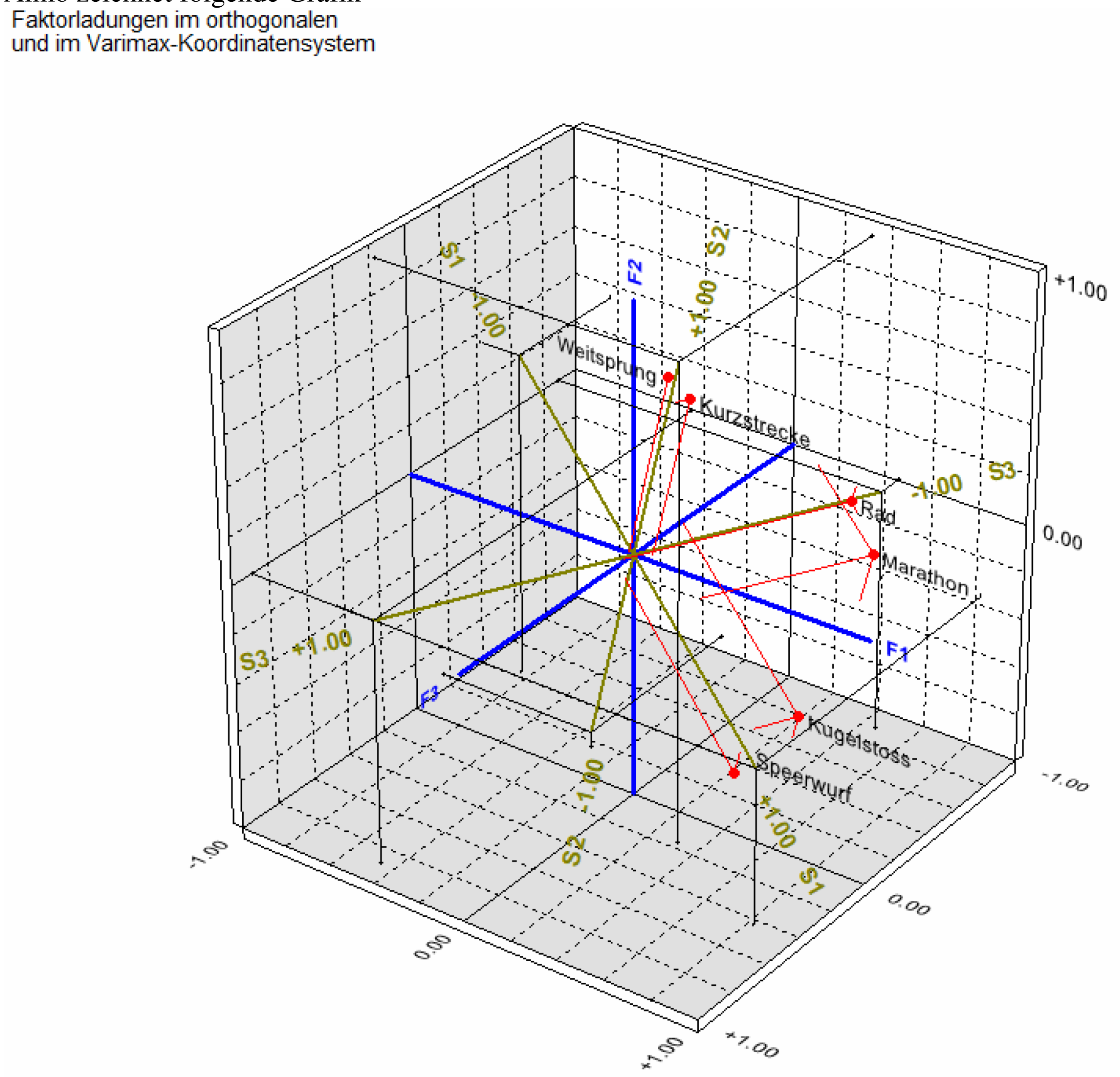
P30.2.3.6 Rechtwinklige Rotation mit 3 Faktoren

Will der Benutzer rechtwinklig rotieren, dann muss er Prog30m2 verwenden. Im einfachen Prog30m1 ist das nicht möglich. Wir wollen aber schon hier die Ausgabe dieser Rotationsart zeigen und erläutern. Almo gibt folgende Ladungsmatrix aus:

```
Rechtwinklig Varimax-rotierte Faktorladungen
```

		Faktor 1	Faktor 2	Faktor 3
Marathon	V1	0.4420	0.2551	-0.6830
Rad	V2	-0.0584	-0.0913	-0.9257
Kugelsto	V3	0.9252	0.1207	-0.1719
Speerwur	V4	0.9180	-0.1194	0.0215
Weitspru	V5	-0.0159	0.9061	0.0158
Kurzstre	V6	0.0295	0.8237	-0.0606

Almo zeichnet folgende Grafik

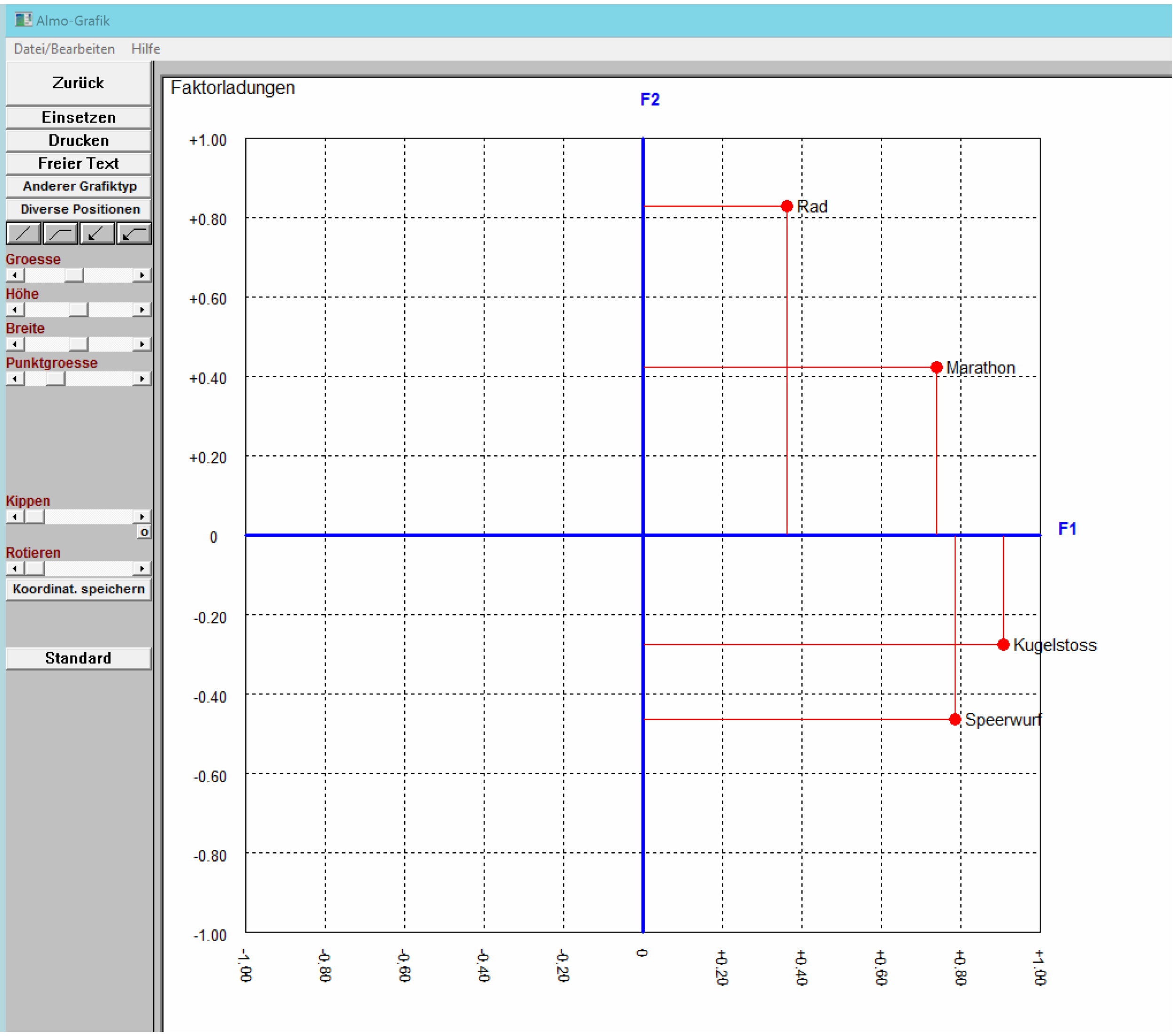


Auch diese Grafik stellt hohe Anforderungen an das räumliche Vorstellungsvermögen des Almo-Benutzers (!).

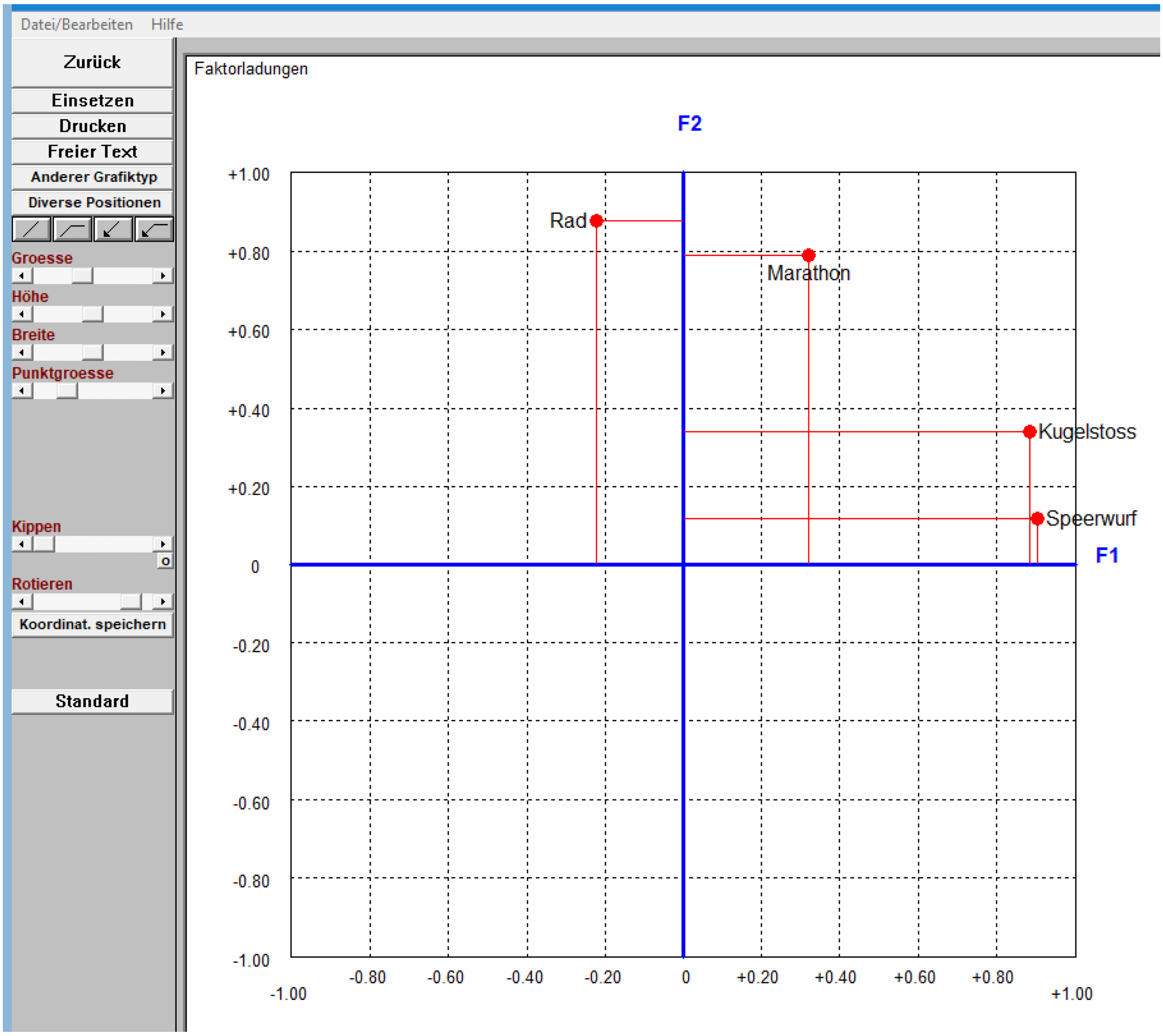
P30.2.3.7 Benutzerdefinierte rechtwinklige Rotation mit 2 und 3 Faktoren

Almo ermöglicht es dem Benutzer, die Punktekonfiguration im (stehen bleibenden) Koordinatensystem zu drehen, d.h. rechtwinklig zu rotieren. Die Vorgehensweise ist folgende

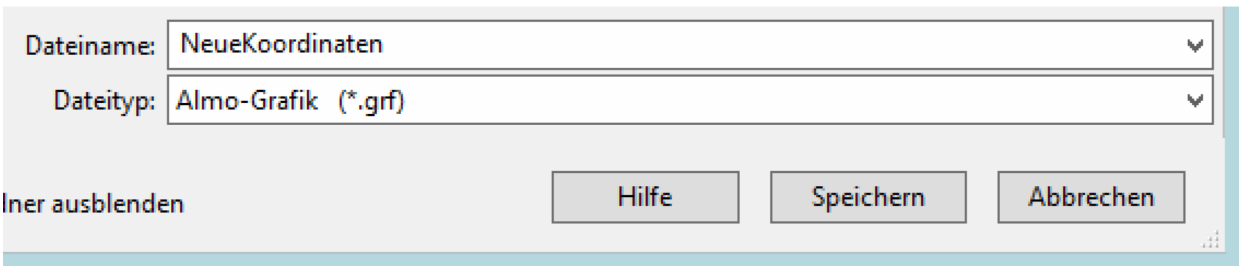
1. Der Benutzer klickt oberhalb der Grafik der unrotierten Faktorladungen auf den großen Grafikknopf. Daraufhin öffnet Almo den Grafik-Editor. Der Benutzer sieht folgendes.



Dies ist nur ein Ausschnitt aus dem Grafik-Editor. In der Werkzeugleiste links befindet sich der Schieber *Rotieren*, darunter der Knopf *Koordinat.speichern*. Wenn der Benutzer den Schieber betätigt, dann beginnt die Punktekonfiguration zu rotieren. Beispielweise wird in folgende Position rotiert:



Die Punktekonfiguration behält beim Rotieren ihre Gestalt bei. Die Distanzen zwischen den Punkten bleiben unverändert.
Wird auf den Knopf *Koordinat.speichern* geklickt, dann öffnet Windows die Datei-Auswahlbox. Zuerst muss ein Ordner festgelegt werden, in den die Dateien gespeichert werden sollen. Almo wird vermutlich den Almo-Unterordner "Progs" vorschlagen. Im unteren Teil der Box sieht man folgendes:



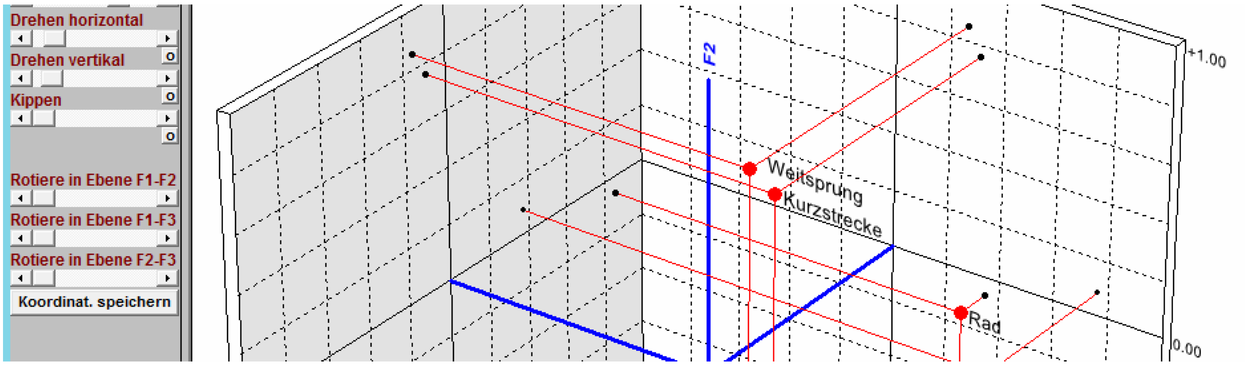
Man gibt einen Dateinamen ohne Punkt und Erweiterung an - im Beispiel: *NeueKoordinaten*, dann Klick auf auf den Knopf *Speichern*. Almo speichert dann

1. eine Datei "NeueKoordinaten.grf", die die Grafik enthält und
2. eine Datei "NeueKoordinaten.mat", die die neue rechtwinklig rotierte Faktor-ladungsmatrix enthält.

Die neue Faktorladungsmatrix in unserem Beispiel ist:

```
     Marathon      0.2928      0.7997
          Rad     -0.2542      0.8667
   Kugelstoss      0.8709      0.3715
    Speerwurf      0.8989      0.1489
```

Bei 3 Faktoren hat die Werkzeugleiste auf der linken Seie des Grafik-Editors folgendes Aussehen. Wir zeigen nur einen Ausschnitt:



Werden die 3 Schieber *Drehen horizontal*, *Drehen vertikal* und *Kippen* betätigt, dann dreht sich der gesamte Würfel. Der Zweck dieser Drehungsart ist es einen guten Einblick in die Punktekonfiguration zu erzielen. Eine Rotation im Sinne der Faktorenanalyse findet nicht statt.

Anders bei den 3 Rotiere-Schiebern. Hier bleiben der Würfel und die Koordinatenachsen unbeweglich stehen. Innerhalb dieser wird die Punktekonfiguration gedreht. Das ist dann "Rotation" im eigentlichen Sinne der Faktorenanalyse. Die Punktekonfiguration bleibt während des Rotierens unverändert, d.h. die Distanzen zwischen den Punkten bleiben gleich.

Rotiert wird separat in 3 Ebenen. Wird beispielsweise in der Ebene F1-F2 rotiert, dann kann man die Drehbewegung gut auf der Rückwand erkennen. Die Reihenfolge, mit der rotiert wird ist beliebig.

Wird auf den Knopf *Koordinat.speichern* geklickt, dann wird wie oben beschrieben Grafik und rotierte Faktorladungsmatrix gespeichert.

P30.3 Eingabe mit vielen Optionen

Die oben angegebene Voreinstellungen können durch eine Fülle von Optionen abgewandelt und erweitert werden. Wir werden das im folgenden ausführen.

P30.3.1 Programm-Maske Prog 30m2

Sie finden dieses Programm durch Klick auf den Knopf „Verfahren“ und dann auf „Faktorenanalyse“.

Prog30m2.Msk
Faktorenanalyse
mit Optionen

Beispiel: Es soll festgestellt werden, ob 6 verschiedene sportliche Leistungen auf einer bzw. mehreren Fähigkeiten (=Faktoren) beruhen.

Eine Standard-Faktorenanalyse würde so verlaufen:

Aus den Variablenwerten der Untersuchungspersonen wird zuerst eine Korrelationsmatrix gebildet. Diese wird dann mit der Hauptkomponenten-Methode (d.h. mit 1.0 in der Diagonalen) faktorisiert. Es entsteht die unrotierte Faktorladungsmatrix. Diese kann als 2- bzw. 3-dimensionales Koordinatensystem grafisch dargestellt werden. Dann wird schiefwinklig rotiert. D.h. es wird versucht, räumlich voneinander getrennte "Punktewolken" zu identifizieren und durch diese Achsen zu ziehen. Diese Achsen können dann als die geometrischen Repräsentanten der (in unserem Beispiel) Fähigkeiten betrachtet werden, die hinter den 6 sportlichen Leistungen stehen. In unserem Beispiel könnten diese mit "Kraft", "Schnelligkeit" und "Ausdauer" benannt werden

Über eine Vielzahl von Optionen, kann diese Standard-Vorgehensweise modifiziert werden. Kurz und schlagwortartig:

Faktorenanalyse als Hauptachsen-Lösung, Hauptkomponenten-Lösung, Alpha-Faktorenanalyse, Image-Faktorenanalyse, kanonische Faktorenanalyse.

Faktorenanalyse angewandt auf die Korrelations- oder Kovarianzmatrix oder auf die Matrix der Abweichungsquadrate

Rechtwinklige Varimax-Rotation oder schiefwinklige Quartimin-Rotation.

Kommunalitätenschätzung: Multiple Bestimmtheitsmasse, oder 1.0 oder größte Korrelation oder selbst gewählte Werte. Iteration der Kommunalitäten.

Faktoren-Signifikanz kann geprüft werden. Festlegen der Faktorenzahl. Ermitteln von Faktorwert-Koeffizienten. Berechnen der euklidischen Distanzen zwischen den Variablen

Siehe Handbuch, Abschnitt P30

Programm-Bedienung ---> Hilfe

Speicher fuer x Variable | Hilfe

Vereinbare Variable= 23

Option: Weitere Vereinbarungen - nur wenn Almo dazu auffordert

Variablennamen

Datei der Variablennamen — Hilfe

".\Testdat\Faktanal.nam"

zeige — zeige = Namensdatei in Output zeigen
leer = nicht zeigen

Freie Namensfelder — Hilfe

Leere alle Eingabefelder dieser Sub-Box

Name1=Marathon
Name2=Rad
Name3=Kugelstoss
Name4=Speerwurf
Name5=Weitsprung
Name6=Kurzstrecke

erzeuge zusätzliche Namensfelder

Variablennamen in Datei speichern — Hilfe
Eingabefeld leer = nicht speichern

Datei aus der gelesen wird — Hilfe
bei Datei-Problemen

".\Testdat\Fakdat.fre"

frei — Format der Daten — Hilfe

V1:6 — der Datensatz enthält diese Variablen
Bei Format DIREKT schreiben Sie: alle_V

Wenn Dateiformat FIX oder Nicht-Standard-FREI — Hilfe

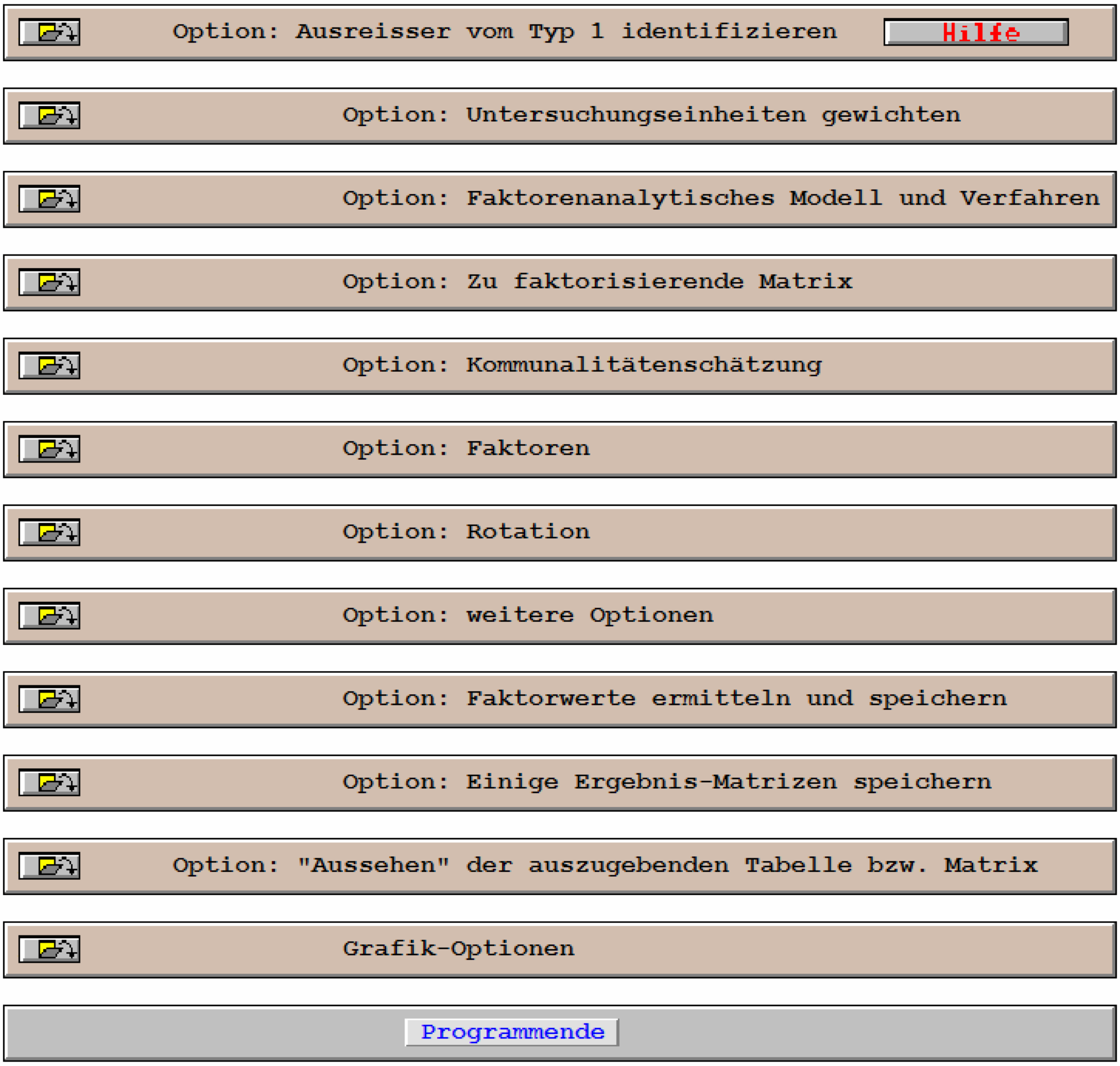
zu faktorisierende quantitative Variable

Marathon,Rad,Kugelstoss,Speerwurf,Weitsprung,Kurzstrecke

Option: Ein- und Ausschliessen von Untersuchungseinheiten

Option: Umkodierungen und Kein-Wert-Angaben

Option: Spezielle Kein-Wert-Behandlung



Erläuterungen zu den Boxen des Maskenprogramms Prog30m2

Eingabebox: Speicher und **Vereinbarungen**
Siehe dazu Almo-Dokument Nr. 0 "Arbeiten mit Almo", Abschnitt P0.1 und P0.2.

Eingabebox: Variablennamen
Siehe dazu Almo-Dokument Nr. 0 "Arbeiten mit Almo", Abschnitt P0.3.

Eingabebox: Datei und **Wenn Dateiformat ...**
Siehe dazu Almo-Dokument Nr. 0 "Arbeiten mit Almo", Abschnitt P0.4.

Eingabebox: Zu faktorisierende quantitative Variable
Siehe P0.11.

Eingabebox: Ein- und Ausschliessen von Untersuchungseinheiten
Siehe P0.7.

Eingabebox: Umkodierungen und Kein-Wert-Angaben
Siehe P0.5.

P30.3.2 Eingabebox: Spezielle Kein-Wert-Behandlung

Option: Spezielle Kein-Wert-Behandlung

Optionsbox geöffnet:

Loesche wieder diese Box

Option: Spezielle Kein-Wert-Behandlung

1

0= Kein-Wert-Fälle in Analyse-Variable nicht vorhanden

1= Paarweises Ausscheiden I <---- ist Voreinstellung

2= Paarweises Ausscheiden II
a. Paarweises Ausscheiden bei quantitativen und ordinalen Variablen.
b. Vollständiges Ausscheiden bei nominalen Variablen und deren Interaktionen, wenn auch nur eine der nominalen Analyse-Variablen den Wert "Kein_Wert" besitzt

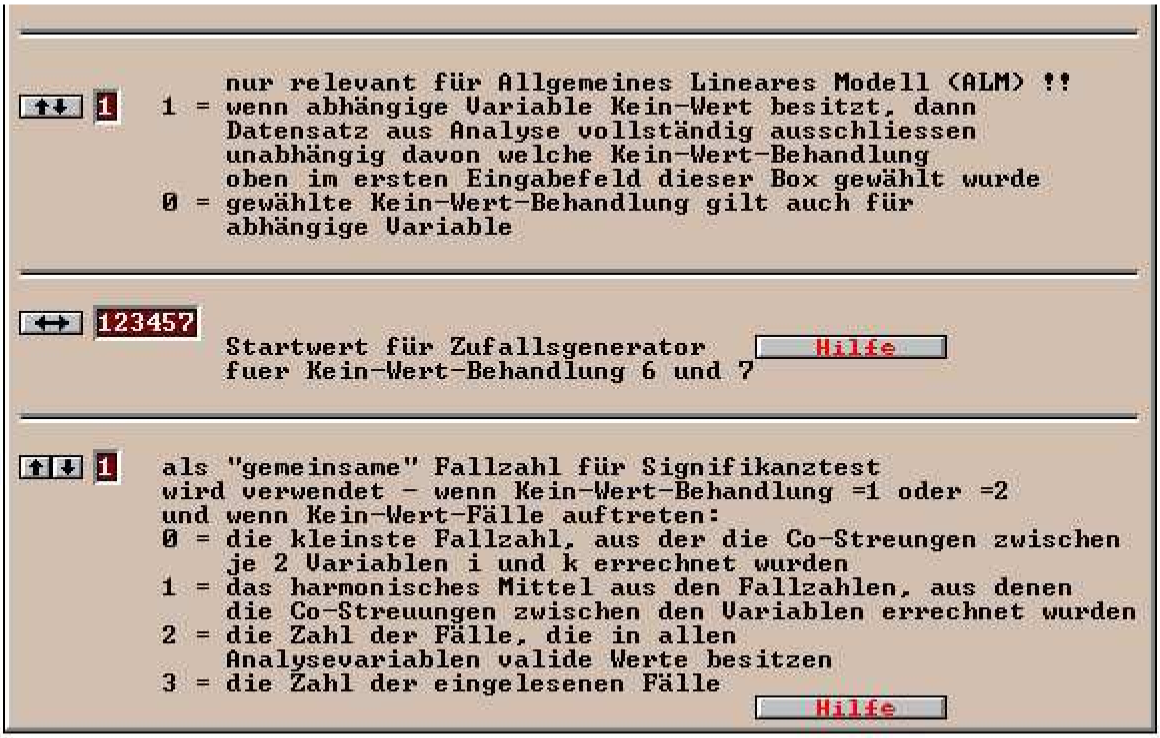
3= Vollständiges Ausscheiden
Vollständiges Ausscheiden des gesamten Datensatzes, wenn auch nur eine der Analyse-Variable "Kein_Wert" ist

4= Mittelwert-Einsetzung I   Hilfe
Für Kein_Wert wird eingesetzt:
a. bei quantitativen Variablen der Mittelwert
b. bei ordinalen Variablen der Median.
Der zum Median nächst gelegene empirische Skalenwert wird dann eingesetzt
c. bei nominalen Variablen der Erwartungswert

5= Mittelwert-Einsetzung II   Hilfe
Für Kein_Wert wird eingesetzt:
a. bei quantitativen Variablen der zum Mittelwert nächste empirisch vorkommende Wert
b. bei ordinalen der Median (wie bei 4)
c. bei nominalen Variablen der Erwartungswert (wie bei 4)

6= Mittelwert-Einsetzung III   Hilfe
Für Kein_Wert wird eingesetzt:
a. bei quantitativen Variablen der Mittelwert +/- einem normalverteilten Zufallswert mit Mittelwert=0 und Standardabweichung der Variablen
b. bei ordinalen der Median (wie bei 4)
Ist die Variable mit gleicher Schrittweite kodiert, dann wird ein Wert X errechnet, der sich ergibt aus Median +/- einem normalverteilten Zufallswert mit Mittelwert=0 und Standardabweichung in der Größe des halben Quartilsabstands der Variablen. Der zu X nächst gelegene empirische Skalenwert wird dann eingesetzt
c. bei nominalen der wahrscheinlichste Ausprägungswert

7= Mittelwert-Einsetzung IV   Hilfe
a. bei quantitativen Variablen zunächst wie bei 6
Der nächst gelegene empirische Skalenwert wird dann eingesetzt
b. bei ordinalen der Median (wie bei 6)
c. bei nominalen Variablen (wie bei 6)



Almo bietet 7 Verfahren an, mit fehlenden Werten umzugehen

0 Kein-Wert-Fälle in Analyse-Variable nicht vorhanden

1 Paarweises Ausscheiden I - ist Voreinstellung

2 Paarweises Ausscheiden II

a. Paarweises Ausscheiden bei quantitativen und ordinalen Variablen.
b. Vollständiges Ausscheiden bei nominalen Variablen und deren Interaktionen, wenn auch nur eine der nominalen Analyse-Variablen den Wert "Kein_Wert" besitzt

3 Vollständiges Ausscheiden

Vollständiges Ausscheiden des gesamten Datensatzes, wenn auch nur eine der Analyse-Variable "Kein_Wert" ist

4 Mittelwert-Einsetzung I

Almo ermittelt zuerst Mittelwerte (für quantiative Variable), Median (für ordinale Variable) und den Erwartungswert (für nominale Variable).

Für Kein_Wert wird dann eingesetzt:

a. bei quantitativen Variablen der Mittelwert
b. bei ordinalen Variablen der Median (=der mittlere Wert). Liegt der Median nicht auf einem empirischen Wert, sondern zwischen 2 empirschen Werten, dann wird der nächst gelegene Nachbarwert als KW-Einsetzungswert verwendet.
c. bei nominalen Variablen die zum Erwartungswert nächste empirisch vorkommende Codeziffer

Die Berechnung des Erwartungswerts soll an einem Beispiel gezeigt werden. Die nominale Variable sei der Beruf mit den 3 Ausprägungen Arbeiter, Angestellte, Sonstige. Dabei wurden folgende Häufigkeiten ermittelt.

	Code	Häufigkeit	Anteil	Code*Anteil
Arbeiter	1	250	0.25	0.25
Angestellte	2	400	0.40	0.80
Sonstige	3	350	0.35	1.05

Summe				2.10

Der Erwartungswert ist 2.1
Die nächste empirisch vorkommende Codeziffer ist 2
Der KW-Einsetzungswert ist also 2

Ist die nominale Variable dichotom, dann ist der Kein-Wert-Einsetzungswert gleich der Codeziffer der häufigsten Ausprägung

5 *Mittelwert-Einsetzung II*

Für Kein_Wert wird eingesetzt:

a. bei quantitativen Variablen der zum Mittelwert nächste empirsch vorkommende Skalenwert
b. bei ordinalen Variablen der Median (=der mittlere Wert). Liegt dieser zwischen zwei empirisch vorkommenden Werten, dann wird der nächst gelegene empirische Nachbarwert als KW-Einsetzungswert verwendet.
c. bei nominalen Variablen die zum Erwartungswert nächste empirisch vorkommende Codeziffer.
   Zur Berechnung des Erwartungswerts siehe oben 4c.

6 *Mittelwert-Einsetzung III*

Für Kein_Wert wird eingesetzt:

a. bei quantitativen Variablen der Mittelwert +/- einem normalverteilten Zufallswert mit Mittelwert=0 und Standardabweichung der Variablen.
b. bei ordinalen Variablen der Median.

   Ist die Variable (was eher ungewöhnlich ist) mit ungleichen Schrittweiten kodiert, dann wird der Median eingesetzt. Liegt dieser zwischen zwei empirisch vorkommenden Werten, dann wird der zum Median nächst gelegene empirische Wert verwendet.

   Ist die ordinale Variable mit gleicher Schrittweite kodiert, dann wird ein Wert X errechnet, der sich ergibt aus Median +/- einem normalverteilten Zufallswert mit Mittelwert=0 und Standardabweichung in der Größe des halben Quartilsabstands der Variablen. Der zu X nächst gelegene empirische Skalenwert wird dann eingesetzt.

   Bei quantitativen und bei ordinalen Variablen wird eine normalverteilte Zufallszahl mit Mittelwert=0 generiert. Als Standardabweichung wird bei quantitativen Variablen die der jeweiligen Variablen verwendet. Bei ordinalen Variablen wird der halbe Quartilsabstand verwendet.

   Betrachten wir ein Beispiel: Die quantitative Variable sei das Lebensalter. Almo errechnet für sie einen Mittelwert von 40. Die Standardabweichung sei 20. Dann wird eine normalverteilte Zufallszahl mit Mittelwert=0 und Standardabweichung=20 erzeugt. Nehmen wir an es entsteht der Zufallswert Z= -15.25. Für den fehlenden Wert wird dann eingesetzt X = 40-15.25 = 24.75.

Bei einer ordinalen Variablen wird entsprechend verfahren. Als Standardabweichung für die Generierung der Zufallszahl wird der halbe Quartilsabstand verwendet. Der ermittelte X-Wert wird bei der ordinalen Variablen aber noch nicht als KW-Einsetzungswert verwendet. Es wird nach dem empirisch vorkommenden Wert gesucht, der am dichtesten bei X liegt. Dieser wird als KW-Einsetzungswert verwendet. So wird verhindert, dass KW-Einsetzungswerte entstehen, die empirisch nicht vorkommen.

c. Bei nominalen Variablen wird der wahrscheinlichste Ausprägungswert eingesetzt.
Die Vorgehensweise soll an einem Beispiel gezeigt werden. Die nominale Variable sei der Beruf mit den 3 Ausprägungen Arbeiter, Angestellte, Sonstige. Dabei wurden folgende Häufigkeiten ermittelt.

	Code	Häufigkeit	in %	in % kummuliert
Arbeiter	1	250	25	25
Angestellte	2	400	40	65
Sonstige	3	350	35	100

Dann wird eine gleichverteilte Zufallszahl zwischen 0 und 100 erzeugt.

Liegt sie zwischen
0 und 25, dann wird für den fehlenden Wert 1 eingesetzt
25 65 2
65 100 3

7 *Mittelwert-Einsetzung IV*
Für Kein_Wert wird eingesetzt:

a. Bei quantitativen Variablen:
Es wird zunächst ein Wert X errechnet, der sich ergibt aus dem Mittelwert +/- einem normalverteilten Zufallswert mit Mittelwert=0 und der Standardabweichung der Variablen. Dann wird der zu X nächst gelegene empirische Skalenwert für Kein_Wert eingesetzt

b. bei ordinalen Variablen der Median.

Ist die Variable (was eher ungewöhnlich ist) mit ungleichen Schrittweiten kodiert, dann wird der Median eingesetzt. Liegt dieser zwischen zwei empirisch vorkommenden Werten, dann wird der zum Median nächst gelegene empirische Wert verwendet.

Ist die Variable mit gleicher Schrittweite kodiert, dann wird ein Wert X errechnet, der sich ergibt aus Median +/- einem normalverteilten Zufallswert mit Mittelwert=0 und Standardabweichung in der Größe des halben Quartilsabstands der Variablen. Der zu X nächst gelegene empirische Skalenwert wird dann eingesetzt.

Die Besonderheit dieser Kein-Wert-Behandlung ist also:

1. Bei quantitativen und bei ordinalen Variablen wird eine normalverteilte Zufallszahl mit Mittelwert=0 generiert. Als Standardabweichung wird bei quantitativen Variablen die der jeweiligen Variablen verwendet. Bei ordinalen Variablen wird der halbe Quartilsabstand verwendet.
Betrachten wir ein Beispiel: Die quantitative Variable sei das Lebensalter. Almo errechnet für sie einen Mittelwert von 40. Die Standardabweichung sei 20. Dann wird eine normalverteilte Zufallszahl mit Mittelwert=0 und

Standardabweichung=20 erzeugt. Nehmen wir an es entsteht der Zufallswert Z= -15.25. X ist dann = 40-15.25 = 24.75.
2. Bei einer ordinalen Variablen wird entsprechend verfahren Als Standardabweichung für die Generierung der Zufallszahl wird der halbe Quartilsabstand der Variablen verwendet.
3. Der ermittelte X-Wert wird bei der quantitativen und der ordinalen Variablen aber noch nicht als KW-Einsetzungswert verwendet. Es wird nach dem empirisch vorkommenden Wert gesucht, der am nächsten bei X liegt. Dieser wird als KW-Einsetzungswert verwendet. So wird verhindert, dass KW-Einsetzungswerte entstehen, die empirisch nicht vorkommen.

c. Bei nominalen Variablen wird der wahrscheinlichste Ausprägungswert eingesetzt. Zur Vorgehensweise siehe oben 6c.

P30.3.3 Eingabebox: Ausreisser identifizieren

```
Option: Ausreisser vom Typ 1 identifizieren          Hilfe
```

Optionsbox geöffnet

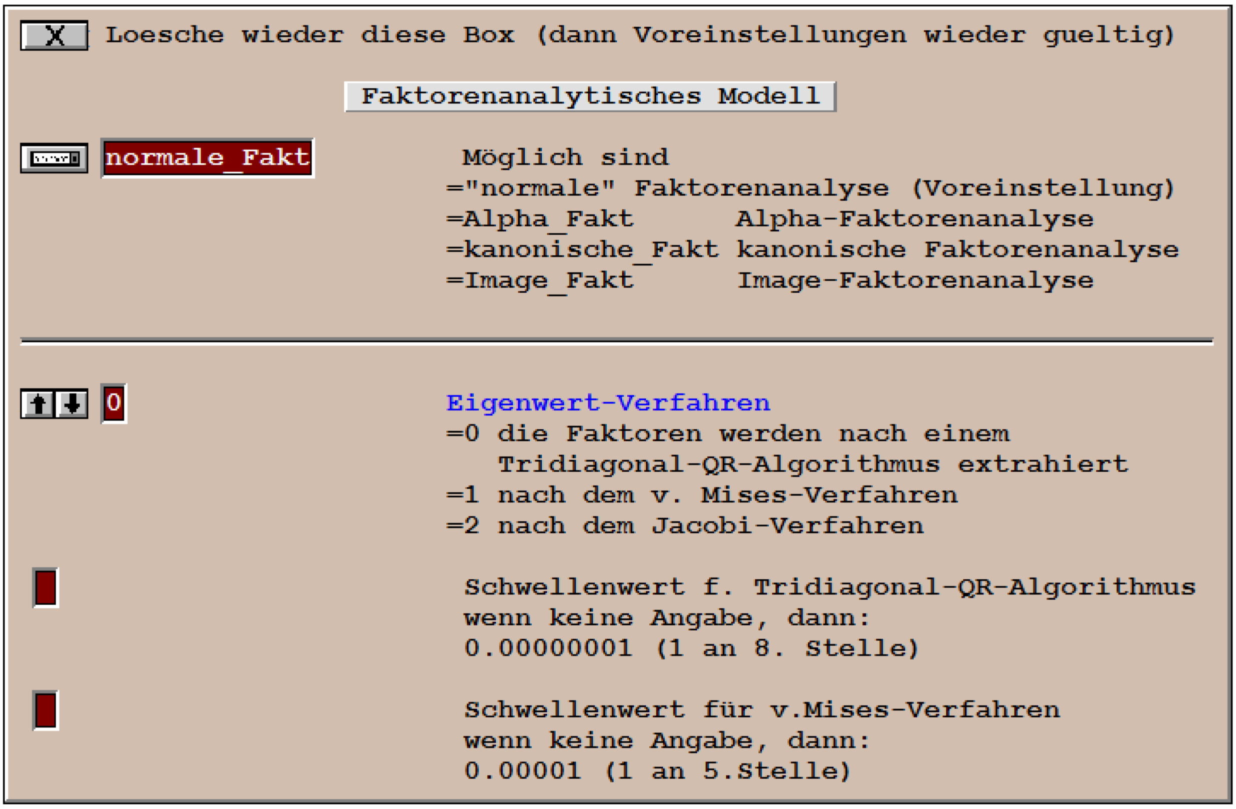
```
X  Loesche wieder diese Box (dann Voreinstellungen wieder gueltig)

          Ausreisser vom Typ 1 identifizieren
        für ordinale und quantitative Analysevariable

                    Was sind Ausreisser ? --->  Hilfe
                     Ausreisser vom Typ 1 --->  Hilfe
                     Ausreisser vom Typ 2 --->  Hilfe

   1      1 = Ausreisser über Quartile
              und Quartilsdifferenz identifizieren
          2 = Ausreisser über Mittelwert
              und Standardabweichung identifizieren
          0 = nicht

   2.5    Multiplikator für Quartilsdifferenz bzw. Stand.abweichung
          Bestimmt Breite des validen Bereichs         Hilfe

   1      Reaktion                                     Hilfe
          0 = nichts tun
          1 = Ausreisser in Ergebnisliste nur melden
          2 = melden u. auf KeinWert setzen
          3 = melden u. auf validen Grenzwert setzen
          4 = melden u. auf bereinigte Ober- bzw. Untergrenze setzen
          5 = melden u. ganzen Datensatz ausschliessen
-----------------------------------------------------------------
                                                       Hilfe
   1      ordinale Variable in Ausreisser-Suche einbeziehen
          1 = einbeziehen
          0 = ausschliessen
              für ordinale Variable nicht nach Ausreissern auchen
```

Im Almo-Dokument Nr 23 "Ausreisser entdecken" wird ausführlich dargestellt, wie Ausreisser in Almo behandelt werden.

Eingabebox: Untersuchungseinheiten gewichten
Siehe P0.8.

P30.3.4 Eingabebox: Faktorenanalytisches Modell und Verfahren

Optionsbox geöffnet:



Eingabefeld 1:

P30.3.4.1 Alpha-, Image-Faktorenanalyse, kanonische Faktorenanalyse

Neben der "normalen" Faktorenanalyse können in Almo noch

die Alpha-Faktorenanalyse
die kanonische Faktorenanalyse
die Image-Faktorenanalyse

gerechnet werden. Die 3 Verfahren werden im Anhang 1 ausführlich beschrieben. Im Anhang A2.6 werden mit Almo und SPSS verschiedene Beispiele gerechnet. Wir empfehlen, sich diesen Anhang anzuschauen.

Diese Faktorenanalyse-Modelle sind ausführlich beschrieben in Holm (1976). Siehe dazu auch Arminger (1979) und Harman (1977).

Anmerkung

1. Als Kommunalitätenschätzungen werden *standardmäßig* eingesetzt

   a. bei der "normalen" Faktorenanalyse 1.0
      Der Benutzer kann, wenn er die Optionsbox "Kommunalitätenschätzung" öffnet, diese Eingabe verändern

b. bei der Image-Faktorenanalyse werden die multiplen Bestimmtheitsmaße (R-Quadrat) als Kommunalitätenschätzung eingesetzt. Der Benutzer kann diese Kommunalitätenschätzung nicht beeinflussen
c. bei der Alpha-Faktorenanalyse 1.0
Der Benutzer kann auch die die multiplen Bestimmtheitsmaße (R-Quadrat) als Kommunalitätenschätzung einsetzen
d. bei der kanonischen und der Image-Faktorenanalyse die multiplen Bestimmtheitsmaße (R-Quadrat)
Der Benutzer kann diese Kommunalitätenschätzung nicht beeinflussen

2. Bei der Alpha-Faktorenanalyse und der kanonischen Faktorenanalyse werden automatisch 25 Kommunalitäten-Iterationen durchgeführt - es sei denn, der Benutzer öffnet die Optionsbox "Kommunalitätenschätzung" und legt eine andere Zahl fest.

3. Die Zahl der zu extrahierenden Faktoren wird so festgelegt
   a. bei der Image-Faktorenanalyse wird die Zahl der Eigenwerte > 1.0 aus der Korrelationsmatrix mit Diagonalwerten 1.0 als Faktorenzahl übernommen.
   Der Benutzer kann diese Automatik nicht beeinflussen. Er kann jedoch die Optionsbox "Faktoren" öffnen und eine von ihm gewünschte Zahl von Faktoren verlangen.
   b. Bei der kanonischen Faktorenanalyse (bei der die Komunalitäten immer iteriert werden) folgen wir einem Vorschlag von Harris (1962, S. 262). Er schlägt vor, für die Iteration jene Faktoren beizubehalten, die bei Iteration 0 eine kanonische Korrelation größer 0 besessen haben. Das sind jene, deren Eigenwerte aus der "Kanon.Fakt.Matrix" in der 1. Analyse größer .0 waren. Siehe dazu Abschnitt A1.1, Formel 3a.
   Der Benutzer kann diese Voreinstellung von Almo durchbrechen. Er muss dazu die Optionsbox " Kommunalitätenschätzung" öffnen, beim Bootstrap-Programm Prog30ml heißt die Optionsbox "Faktoren und Kommunalitäten" Wir zeigen einen zusammen geschnittenen Ausschnitt

```
Kommunalitätenschätzung

R_Quadrat
      Almo setzt in die Diagonale der Matrix als
      geschätzte Kommunalitäten ein:
```

```
      = R_Quadrat
        die multiplen Bestimmtheitsmaße
        1     0 = dabei werden alle Faktoren mit Eigenwert
                  grösser 0 extrahiert
              1 = groesser 1 extrahiert          Hilfe
```

1. Eingabefeld: **R_Quadrat *muss* gewählt werden**

2. Eingabefeld:

**0 = es werden alle Faktoren mit Eigenwert größer 0 extrahiert
Dabei werden die Eigenwerte berechnet aus der Korrelationsmatrix
mit den multiplen Bestimmtheitsmassen in der Diagonalen.**

**1 = es werden alle Faktoren mit Eigenwert größer 1 extrahiert
Dabei werden die Eigenwerte berechnet aus der Korrelationsmatrix
mit 1.0 in der Diagonalen. Danach werden die multiplen Bestimmtheitsmasse**

```
in die Diagonale der Korrelationsmatrix eingesetzt und die
Faktorenanalyse gerechnet
```

Die Zahl der so extrahierten Faktoren kann bei 0 größer sein als bei 1

c. Bei der Alpha-Faktorenanalyse werden jene Faktoren beibehalten, die bei Iteration 0 einen Alpha-Wert > 0 besitzen. Das sind jene, deren Eigenwerte aus der "Alphamatrix" bei der 1. Analyse größer 1.0. waren. Siehe dazu Abschnitt A1.2, Formel 5 und die darauf folgenden Ausführungen.
   Der Benutzer kann diese Voreinstellung von Almo durchbrechen. Er muss dazu ebenfalls die Optionsbox " Kommunalitätenschätzung" öffnen und sich zwischen 0 und 1 entscheiden.

Prinzipiell gilt jedoch, dass der Benutzer die Faktorenzahl selbst festlegen kann. Er muss dazu die Optionsbox "Faktoren" öffnen. Im Bootstrap-Programm 30ml heißt die Optionsbox "Faktoren und Kommunalitäten".

Eingabefeld 2 bis *4:*
Eigenwert-Verfahren (Faktor-Extrahierung)

```
1= die Faktoren werden nach einem
   Tridiagonal-QR-Algorithmus extrahiert
2= nach dem v. Mises-Verfahren
3= nach dem Jacobi-Verfahren
```

In Almo wird standardmäßig ein Tridiagonal-QR-Algorithmus zur Extrahierung der Faktoren verwendet (siehe dazu: Communications of the ACM, Vol.8, Algorithm 254).

Dieser Algorithmus benötigt einen Schwellenwert, der die Präzision der Ergebnisse bestimmt. Dieser wird in Eingabefeld 3 eingetragen.

In Almo ist dieser Schwellenwert auf 0.00000001 (1 an der 8. Stelle) voreingestellt.

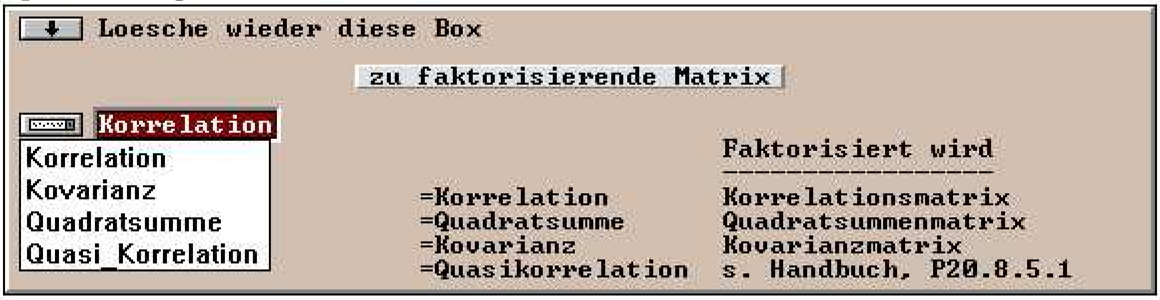
Wird im 3. Eingabefeld z.B. 0.0000000001 geschrieben (1 an der 10. Stelle), dann wird die Genauigkeit der Ergebnisse erhöht. Allerdings wird dies der Benutzer, wenn er sich die Ergebnisse der Faktorenanalyse mit 4 oder 5 Kommastellen ausgeben läßt, kaum bemerken (siehe die Optionsbox 20 „Aussehen der auszugebenden Tabelle“).

Als alternatives Eigenwert-Verfahren ist in Almo das Jacobi-Verfahren und das v. Mises'sche Iterationsverfahren enthalten. Wird im 1. Eingabefeld „1“ eingesetzt, dann kann der Benutzer dieses aktivieren. 0 ist der voreingestellte Tridiagonal-QR-Algorithmus. Das v. Mises-Verfahren ist etwas langsamer. Wenn jedoch bei sehr großen Matrizen nur die ersten 2 oder 3 Faktoren extrahiert werden sollen, dann kann es deutlich schneller sein.

Für das v. Mises'sche Iterationsverfahren kann im 4. Eingabefeld der Schwellenwert vom Benutzer verändert werden. Voreingestellt ist `0.00001` (1 an 5. Stelle) als Schwellenwert. Zum Kalkül des v. Mises'schen Iterationsverfahren siehe Holm, 1975, Kapitel Matrixalgebra, Punkt 38. Beim v. Mises-Verfahren wird ein gewisser Kalkül so oft iteriert, bis die Rayleigh- Quotienten zweier aufeinander folgender Iterationsgänge sich nur noch um 0.00001 unterscheiden. Bei einer klaren Faktorenstruktur sind in der Regel nicht einmal 20 Iterationen dieses Kalküls notwendig. Bei dem in Almo gewählten Voreinstellungen wird (wenn diese Mindestdifferenz nicht früher erreicht wird) nach 100 Iterationen abgebrochen und die bis dahin ermittelte Differenz der zwei Rayleigh-Quotienten unter der Bezeichnung "Toleranz" ausgegeben. Ist dem Benutzer die angegebene "Toleranz" zu groß, dann kann er im „selbst geschriebenen Almo-Programm“ durch **Option 55** die Zahl der Iterationen erhöhen, z.B. auf 150.

P30.3.5 Eingabebox: Zu faktorisierende Matrix

Optionsbox geöffnet:



Die Faktorenanalyse kann nicht nur wie üblich auf die Korrelationsmatrix angewendet werden, sondern auch auf die Quadratsummen- oder Kovarianzmatrix, oder die Quasi-Korrelationsmatrix. Zu letzterer siehe Handbuch zu P45 "Data Mining", Abschnitt P45.12.4.2.

Anmerkungen

0. Der Normalfall wird es sein, die Korrelationsmatrix zu faktorisieren. Man rechnet dann mit standardisierten Variablen, die vergleichbar sind. Eher ungewöhnlich (aber aus wohl überlegten Grunde) wird man die Kovarianzmatrix faktorisieren. Sehr ungewöhnlich ist es, die Quadratsummenmatrix zu faktorisieren.

1. Wenn als Matrix QUADRATSUMME oder KOVARIANZ gesetzt wurde und in der Optionsbox „Kommunalitätenschätzung“ die Diagonale auf 1.0 gesetzt wurde, dann bedeutet dies, dass in die Diagonale die Streuungen der jeweiligen Variablen, also ihre Abweichungs-quadratsummen bzw. Varianzen eingesetzt werden. Es werden alle Faktoren extrahiert, die einen Eigenwert größer 0 besitzen.
   Bei *Kommunalität=R_Quadrat* (siehe nachfolgende Optionsbox) wird für jede Variable die Quadratsumme bzw. die Varianz ermittelt, die durch die jeweils anderen Variablen erklärt wird. Diese wird dann in die Diagonale eingesetzt.
   Bei *Kommunalität=halbeDiag* wird der halbe Wert der in der Diagonale stehenden Streuung als Komunalitätenschätzung verwendet.
   Bei *Kommunalität=maxKorr* wird die größte Streuung in der Spalte (bzw. Zeile) als Komunalitätenschätzung eingesetzt.
   Ein Problem ist es, die Zahl der zu extrahierenden Faktoren zu bestimmen. Almo verwendet hier einen Trick: Die Kovarianzmatrix wird temporär in die Korrelationsmatrix (mit 1.0 in der Diagonalen) überführt und dann die Zahl der Eigenwerte > 1.0 bestimmt. Siehe nach-folgendes Beispiel. Bei der Quadratsummenmatrix funktioniert das nicht. Der Benutzer muss in der Optionsbox "Faktoren" die Zahl der zu extrahierenden Faktoren festlegen.

2. Die Faktorisierung der Quadratsummen-Matrix und der Kovarianzmatrix führen zu proportionalen Faktorladungen (und Faktorwert-Koeffizienten sowie Faktor-werten). Die Faktorladungen aus der Quadratsummenmatrix sind um den Faktor $\sqrt{n}$ (n=Zahl der Untersuchungseinheiten) größer als die aus der Kovarianzmatrix. Bei den Faktorwert-Koeffizienten (die in der Optionsbox „weitere Optionen“

angefordert werden können) und den Faktorwerten (die in der Optionsbox „Faktorenwerte ermitteln und speichern“ angefordert werden können) ist es genau umgekehrt (siehe dazu P30.5 und P30.6). Die Ergebnisse aus der Korrelationsmatrix sind aber immer andere.

Eine Sonderstellung nimmt hier die Kanonische Faktorenanalyse, die Faktorenanalyse der Image-Matrix und die Alpha-Faktorenanalyse ein. Bei diesen Verfahren sind die Faktorladungsmatrizen aus der Kovarianzmatrix **V** und der Korrelationsmatrix **R** zu einander proportional. Wir verwenden folgende Notation:

T = dies sei eine m x m Diagonalmatrix, in deren Diagonale die Kehrwerte der Standardabweichungen der m Variablen (also: $1/\sigma$) enthalten ist.

F = dies ist die Faktorladungsmatrix aus der Korrelationsmatrix R

C = Faktorladungsmatrix aus Kovarianzmatrix V

Für die kanonische, die Alpha-Faktorenanalyse und die skalierte Image-Faktorenanalyse gilt nun

F = **T** * **C**

D.h. die Faktorladungen **c** einer Variablen i (=Zeile i aus **C**) werden mit der Standardabweichung dividiert, um **f** (= Zeile i aus **F**) zu erhalten.

Soll also die Kovarianzmatrix mit Hilfe einer der drei genannten faktorenanalytischen Verfahren faktorisiert werden, dann kann (1) entweder die V-Matrix unmittelbar oder (2) unter Benutzung obiger Gleichung 14 die R-Matrix diesem Verfahren unterworfen werden.

3. Die Faktorisierung der Quadratsummen- bzw. Kovarianzmatrix ist nur dann sinnvoll, wenn die zu faktorisierenden Variablen dieselben Maßeinheiten besitzen. Dies ist z.B. bei Fragebatterien in der Umfrageforschung gegeben, wenn alle Fragen dasselbe vorgegebene Antwortmuster besitzen.

4. Die aus den Faktorwert-Koeffizienten der Quadratsummen- bzw. Kovarianzmatrix ermittelten Faktorwerte sind nicht standardisiert – im Unterschied zu den Faktorwerten aus der Korrelationsmatrix.

Wir rechnen ein Beispiel mit den Daten zu 4 sportlichen Leistungen und stellen die Ergebnisse für die faktorisierte Korrelationsmatrix und Kovarianzmatrix nebeneinander. Als Kommunalitäten wurde in der Optionsbox die R_QUADRAT eingesetzt

```
     Korrelationsmatrix                        Kovarianzmatrix
     ==================                        ===============

Eingesetzte Werte für Kommunalitaeten   Eingesetzte Werte für Kommunalitaeten
(Multiple Bestimmtheitsmasse)            (Multiple Bestimmtheitsmasse)
```

Marathon	0.4362
Rad	0.1782
Kugelsto	0.6828
Speerwur	0.5818

Marathon	356.3309
Rad	174.1361
Kugelsto	439.0030
Speerwur	416.7071

```
-------------------------------------------------------------------------

 Eigenwerte aus Korrelationsmatrix      Eigenwerte groesser 0 aus Kovarianzmatrix
 mit R_Quadrat in der Diagonalen        mit Varianzen in der Diagonalen
```

```
 --------------------------------        1572.46874  997.56938  455.35223  127.80925
 alle Eigenwerte > 0
     1.6856      0.5031                  Um die Zahl der zu extrahierenden Faktoren
                                         zu bestimmen, wird die Kovarianzmatrix
                                         temporaer in die Korrelationsmatrix mit 1.0
                                         in der Diagonalen ueberfuehrt.
                                         2 Eigenwerte der Korrelationsmatrix sind
                                         groesser 1.0.
                                         Entsprechend versucht Almo 2 Faktoren fuer
                                         die Faktorenanalyse der Kovarianzmatrix
                                         zu extrahieren

Prozent der Varianz je Faktor            Prozent der Varianz je Faktor
   42.1401     12.5777                      38.1721     13.4569

Zu erklaerende Gesamtvarianz=4.0000      Zu erklaerende Gesamtvarianz=3153.1996
Durch 2 Faktoren                         Durch 2 Faktoren
erklaerte Varianz=            2.1887     erklaerte Varianz=         1627.9643
Prozentsatz der erklaerten               Prozentsatz der erklaerten
Varianz=                     54.7177     Varianz=                     51.6290

--------------------------------------------------------------------------------------
```

Matrix der Faktorladungen

	Faktor 1	Faktor 2
Marathon	0.5875	0.4227
Rad	0.2244	0.4491
Kugelsto	0.8713	-0.1279
Speerwur	0.7286	-0.3261

Matrix der Faktorladungen

	Faktor 1	Faktor 2
Marathon	17.5756	11.0601
Rad	8.1225	13.1383
Kugelsto	21.9198	-5.0655
Speerwur	18.6624	-10.1845

P30.3.6 Eingabebox: Kommunalitätenschätzung

Option: Kommunalitätenschätzung

Optionsbox geöffnet:

X Loesche wieder diese Box (dann Voreinstellungen wieder gueltig)

Kommunalitätenschätzung

1.0

Almo setzt in die Diagonale der Matrix als
geschätzte Kommunalitäten ein:

= 0
Diagonale wird vom Benutzer eingegeben
(Eingabe der Werte siehe weiter unten)
Ist nicht möglich bei Prog30m4, Prog30ml Hilfe

= 1.0
in die Diagonale wird 1.0 eingesetzt
alle Faktoren mit Eigenwert grösser 1.0 werden extrahiert
(ist Voreinstellung bei normale_ u. Alpha_Fakt.analyse

= R_Quadrat
die multiplen Bestimmtheitsmaße
0 0 = dabei werden alle Faktoren mit Eigenwert
grösser 0 extrahiert
1 = groesser 1 extrahiert Hilfe

= max_Korr
In die Diagonale wird die maximale Korre-
lation (aus Spalte) eingesetzt. Hilfe
Faktoren mit Eigenwert grösser 1.0 werden extrahiert

= halbe_Diag
In die Diagonale wird die Hälfte des
Diagonalwerts eingesetzt
(bei Korrelationsmatrix also 0.5) Hilfe
Faktoren mit Eigenwert grösser 1.0 werden extrahiert

BEACHTE: Wird nicht die Korrelationsmatrix
faktorisiert, dann haben diese Parameter
teilweise eine andere Bedeutung (s.Handbuch)

Selbst eingegebene Kommunalitäten
nicht möglich bei eingegebener fertigen Korrelationsmatrix
bei Prog30m4 und Bootstrap mit Prog30ml. Hilfe

falls oben "0" als Kommunalitätenschätzung eingegeben
wurde, dann müssen Sie hier die Werte eingeben werden
etwa so:
0.8, 0.7, 0.8, 0.7, 0.7, 0.5

BEACHTE: Punkt als Dezimalzeichen
Komma als Trennzeichen
zum Schluss kein Komma mehr

Zahl der Kommunalitäten-Iterationen
0= keine Iteration
x= x Iterationen

wenn keine Angabe, dann
0 Iterationen bei normale und Image_Fakt
10 Iterationen bei Alpha- und kanon. Fakt

Schwellenwert für Kommunalitäten-Iteration
wenn keine Angabe, dann: 0.001

Voreinstellung
Bleibt die Optionsbox „Kommunalitätenschätzung“ geschlossen, dann ist bei der "normalen" Faktorenanalyse 1.0 voreingestellt.

Eingabefeld 1:

Möglich sind folgende DIAGONALE-Angaben:

1	In die Diagonale wird 1.0 eingesetzt. Dies ist die Voreinstellung, braucht also nicht geschrieben zu werden.
R_Quadrat	Es werden die R^2, also die multiplen Bestimmtheitsmaße in die Diagonale eingesetzt
maxKorr	Es wird die maximale Korrelation in der Zeile i (bzw. Spalte i) gesucht und als Kommunalität für die Variable i eingesetzt.
halbeDiag	In die Diagonale wird 0.5 eingesetzt. Diese etwas seltsam anmutende Kommunalitätenschätzung ist eigentlich nur in Verbindung mit einer Kommunalitäten-Iteration sinnvoll.
=0	In die Diagonale werden vom Benutzer festgelegte Werte eingegeben. Siehe dazu auch Abschnitt P30.4.3.

Spezielle Option (nur) bei R_Quadrat
Wurde oben im 1. Eingabefeld *R_Quadrat* eingegeben, dann kann durch Klick auf den 0-1 -Knopf (unterhalb des Eintrags *R_Quadrat)* gewählt werden, nach welcher Methode die Zahl der zu extrahierenden Faktoren erfolgen soll

```
0 = es werden alle Faktoren mit Eigenwert größer 0 extrahiert
    Dabei werden die Eigenwerte berechnet aus der Korrelationsmatrix
    mit den multiplen Bestimmtheitsmassen in der Diagonalen

1 = es werden alle Faktoren mit Eigenwert größer 1 extrahiert
    Dabei werden die Eigenwerte berechnet aus der Korrelationsmatrix
    mit 1.0 in der Diagonalen
```

Wir empfehlen eher 1. Die Zahl der extrahierten Faktoren kann bei 0 unangemessen groß werden.
Diese Zusatzoption zu R_Quadrat ist nur für die "normale" Faktorenanalyse verfügbar, nicht für die Alpha-, Image- und kanonische Faktorenanalyse.

Eingabefeld 2:
Falls Sie oben im Eingabefeld 1 als Kommunalitätenschätzung "0" eingegeben haben, dann müssen Sie hier die Werte der Kommunalitäten eingeben etwa so:

0.8, 0.7, 0.8, 0.7, 0.7, 0.5

BEACHTE: Komma als Trennzeichen. Zum Schluss kein Komma mehr.

P30.3.6.1 Kommunalitäten-Iteration

***Eingabefeld 3* und *4*:** Kommunalitäten-Iteration und Schwellenwert
Nachdem die (unrotierte) Faktorladungsmatrix errechnet wurde, können die Kommunalitäten ermittelt werden. Sie ergeben sich als Summe der quadrierten Faktorladungen. Wir bezeichnen sie als von der Faktorenanalyse *reproduzierte Kommunalitäten*
Beispiel:

```
                 Faktorladungen        reproduzierte
                                       Kommunalitäten
```

	Faktor 1	Faktor 2	
V1	0.4	0.5	$0.4^2 + 0.5^2 = 0.41$
V2	0.6	0.4	$0.6^2 + 0.4^2 = 0.52$
.	.	.	.
.	.	.	.

Nun ist der Gedanke naheliegend, die so ermittelten Kommunalitäten in eine 2. Faktorenanalyse einzusetzen, deren Ergebnisse dann für die Kommunalitäten-schätzung einer 3. Analyse verwendet werden usw. Diese Kommunalitäten-Iteration wird abgebrochen, wenn sich *eingesetzte* und *reproduzierte* Kommunalitäten um weniger als einen Schwellenwert (von etwa 0.001 absolut) unterscheiden.

Betrachten wir ein Beispiel-Programm. Wir geben an:

```
in
Eingabefeld 1: R_Quadrat (=Start-Kommunalitäten)
Eingabefeld 3: 10 (= Iterationen)
Eingabefeld 4: 0.001 (= Schwellenwert)
```

Zuerst werden die multiplen Bestimmtheitsmaße R^2 als Start-Kommunalitäten eingesetzt. Mit dem Schwellenwert wird festgelegt, dass die Kommunalitäten-Iteration abgebrochen wird, wenn eingesetzte und errechnete Kommunalitäten für die einzelnen Variablen nicht mehr als um 0.001 differieren. Mit der Iterationszahl 10 wird festgelegt, dass aber maximal nicht mehr als 10 mal iteriert werden soll (also 11 Faktorenanalysen gerechnet werden sollen).

Der Schwellenwert ist Almo-intern auf 0.001 voreingestellt (braucht also gar nicht geschrieben zu werden).

Folgendes ist zu beachten:

1. Es muss nicht notwendigerweise zu einer Konvergenz kommen.
2. Es können Kommunalitäten größer 1.0 auftreten. Almo gibt in diesem Falle die vorhergehende Lösung als die endgültige aus.
3. Wird eine 2. iterierte Faktorenanalyse mit anderen Start-Kommunalitäten gerechnet, dann kann durchaus das endgültige Ergebnis verschieden sein - wenn auch in der Regel nur wenig verschieden.

Almo iteriert mit einer konstanten Faktorenzahl. Es ist jedoch möglich, dass im Verlauf der Iterationen ein weiterer Faktor (gemäß dem Kaiser- bzw. Guttman-Kriterium - siehe nachfolgend P30.3.5) hinzukommt. Almo akzeptiert diesen Faktor nicht. Der Benutzer erhält folgende Warnung: "Während der Kommunalitäten-Iteration hat sich die Zahl der validen Faktoren erhöht - sie wird wieder reduziert."

Konvergenz der Kommunalitäten-Iteration

Interessant ist es, zu überprüfen, ob das Iterieren kovergiert, d.h. ob die Kommunalitäten-Differenz je Variable gegen 0 strebt.

Wir rechnen das 3-faktorielle Beispiel aus Abschnitt P30.2.3.4. Die Optionsbox "Kommunalitätenschätzung" wird geöffnet und im 3. Eingabefeld 12 Iterationen angefordert. Almo gibt folgendes Iterationsprotokoll aus:

Zuerst wird mitgeteilt, dass bei der 9. Iteration für eine Variable eine reproduzierte Kommunalität größer 1.0 entstand. Almo beendet das Iterieren deswegen mit 8 Iterationen und bringt folgende Warnung

```
                    ******************** WARNUNG
       Kommunalitaet groesser Diagonalglied (1.0) aufgetreten
bei Kommunalitaet-Iteration Nr. 9. Iterationszahl auf 8 herabgesetzt
```

Folgende Ergebnisse werden mitgeteilt:

Kommunalitaeten-Iterationen: 8
Letzte Iterationsdifferenz je Variable

Marathon	V1	0.0029
Rad	V2	0.0057
Kugelsto	V3	0.0187
Speerwur	V4	0.0148
Weitspru	V5	0.0271
Kurzstre	V6	0.0067

Konvergenz der Kommunalitaeten-Iterationen

Kommunalitaeten-Differenz	Maratho	Rad	Kugelst	Speerwu	Weitspr	Kurzstr
von Iteration -1 zu 0	0.273	0.131	0.100	0.143	0.178	0.317
von Iteration 0 zu 1	0.149	0.048	0.027	0.077	0.048	0.185
von Iteration 1 zu 2	0.072	0.016	0.003	0.046	0.005	0.101
von Iteration 2 zu 3	0.033	0.004	0.015	0.032	0.025	0.052
von Iteration 3 zu 4	0.015	0.002	0.019	0.025	0.031	0.027
von Iteration 4 zu 5	0.007	0.004	0.020	0.021	0.031	0.016
von Iteration 5 zu 6	0.004	0.005	0.020	0.019	0.030	0.011
von Iteration 6 zu 7	0.003	0.005	0.019	0.017	0.029	0.008
von Iteration 7 zu 8	0.003	0.006	0.019	0.015	0.027	0.007
konvergiert?	ja	nein	nein	ja	nein	ja
mit Toleranz von 0.009	ja	ja	nein	ja	nein	ja

Mit den in die Diagonale der Korrelationsmatrix "eingesetzten Kommunalitäten" i wird die Faktorenanalyse i gerechnet. Daraus entstehen die "reproduzierten Kommunalitäten" i+1, die ihrerseits nun in die Diagonale der Korrelationsmatrix für die Faktorenanalyse i+1 eingesetzt werden, woraus die reproduzierten Kommunalitäten i+2 entstehen, die ihrerseits ... usw.

Die in der Konvergenz-Tabelle aufgelisteten Kommunalitäten-Differenzen sind

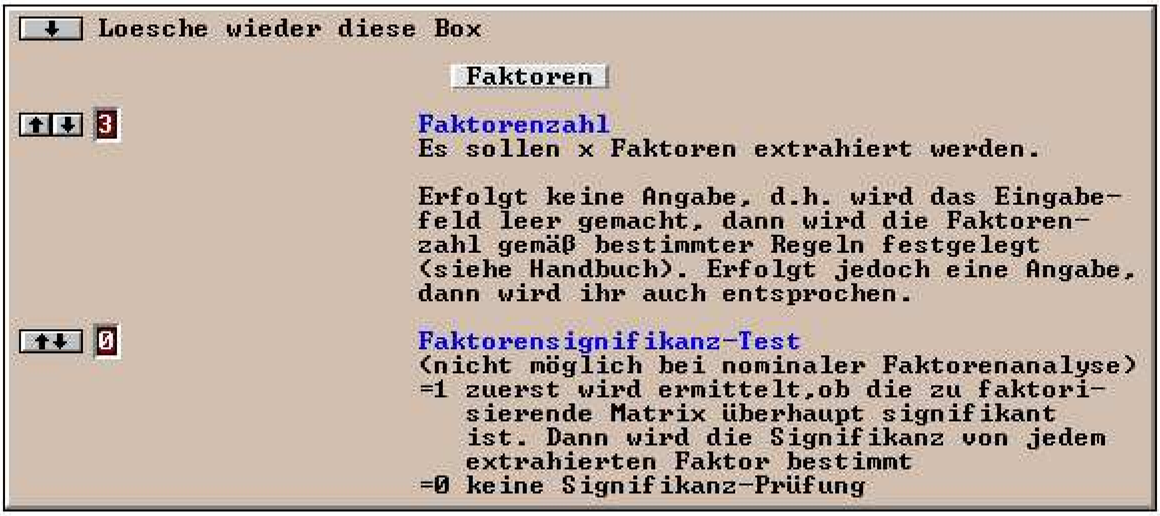
> die absolute Differenz von jeweils "eingesetzter Kommunalität" i minus der
> aus der Faktorenanalyse hervor gegangenen "reproduzierten Kommunalität" i+1
> kurz: die Differenz von eingesetzter minus reproduzierter Kommunalität
> Beachte: Die Differenz ist vorzeichenlos, da sie absolut genommen wird.

Almo prüft bis zur 3. Dezimalstelle, ob die Differenzwerte konvergieren. Das "nein" bei der Variablen "Rad" könnte in ein "ja" gewandelt werden, da die letzten 5 Differenzwerte nur zwischen 0.002 und 0.006 pendeln. Wird mit einem Toleranzwert von +0.009 zum jeweils vorausgehenden Wert geprüft, dann kann man eine angemessene Konvergenz konstatieren. Bei "Speerwurf" und "Kuzstrecke" jedoch nicht. Deren Differenzwerte bewegen sich offensichtlich nicht in Richtung auf 0.

Mit **-1** in der ersten Zeile der Tabelle sind die in die Diagonale der Korrelationsmatrix eingesetzten Anfangs-Kommunalitäten gemeint, die standardmäßig von Almo eingesetzt wurden bzw. vom Benutzer angefordert wurden. In unserem Beispiel ist das 1.0.
Kurz: Die 1. Zeile der Tabelle enthält die vorgegebene minus der reproduzierten Kommunalität.

P30.3.7 Eingabebox: Faktoren

Optionsbox geöffnet:



Eingabefeld 1:

P30.3.7.1 Faktorenzahl (Kaiser- und Guttman-Kriterium)

Almo ist mit folgender Automatik ausgestattet.

Wird in der Optionsbox „Kommunalitätenschätzung“ die Diagonale auf R_QUADRAT gesetzt, dann werden alle Faktoren extrahiert, die einen Eigenwert größer 0 besitzen. Dies ist das "**Guttman-Kriterium**". Der Benutzer kann diese Vorgehensweise modifizieren. Siehe oben "*Spezielle Option (nur) bei R_Quadrat*"

Bei allen anderen DIAGONALE-Angaben verfährt Almo folgendermaßen: Es setzt zuerst 1.0 in die Diagonale und ermittelt die Zahl f der Eigenwerte, die größer 1.0 sind. Dies ist das "**Kaiser-Kriterium**". Dann setzt Almo die Kommunalitäten-schätzungen ein, die der Benutzer durch seine DIAGONALE-Anweisung bestimmt hat und extrahiert f Faktoren. (Faktoren mit Eigenwert kleiner 0 können selbstverständlich nicht ermittelt werden).

Der Benutzer kann diese Automatik durchbrechen, wenn er im 1. Eingabefeld der Optionsbox "Faktoren"eine Zahl einsetzt.

Betrachten wir ein Beispiel: Der Benutzer setzt für die Diagonale „MaxKorr“ ein und in der Box „Faktoren“ im 1. Eingabefeld 10. Auf Grund der Diagonale-Angabe würde das Kaiser-Kriterium wirksam werden. D.h. es würden eigentlich nur jene Faktoren extrahiert werden, die bei einer vorausgehenden Analyse mit 1.0 in der Diagonalen einen Eigenwert größer 1.0 besitzen.
Die Faktoren-Angabe des Benutzers dominiert aber die Diagonale-Anweisung.

Almo wird 10 Faktoren extrahieren. Selbstverständlich können nur Faktoren mit einem Eigenwert größer/gleich 0 extrahiert werden. Der Benutzer kann auch eine geringere Faktorenzahl angeben, als auf Grund des Kaiser-Kriteriums extrahiert würde.

Eingabefeld 2:

P30.3.7.2 Signifikanz der Matrix und der Faktoren

Wird in das 2. Eingabefeld "1" eingesetzt, dann wird (1.) die Signifikanz der zu faktorisierenden Matrix und (2.) die Signifikanz der extrahierten Faktoren ermittelt.
Die zu faktorisierende Matrix wird einem statistischen Signifikanztest (nach Bartlett) unterworfen. Siehe Holm, 1976, Abschnitt 2.1.1. Die Formel ist folgende

Chi-Quadrat = (N - (2m + 5) / 6) * ln(det **R**)
Freiheitsgrade = m * (m -1) / 2

N = Zahl der Untersuchungseinheiten
m = Variablenzahl
ln(det R) = natürlicher Logarithmus der Determinante der Korrelationsmatrix (mit 1.0 in der Diagonale)

Ebenso werden die extrahierten Faktoren einem statistischen Signifikanztest (nach Rippe) unterworfen. Siehe Holm, 1976, Abschnitt 4.2.
Die Formel ist folgende

$$\text{Chi} - \text{Quadrat}_{\text{i}} = (\boldsymbol{N} - 1) * \ln \frac{\det \boldsymbol{Q}_i}{\det \boldsymbol{R}}$$

det Q_i= Determinante der reproduzierten Korrelationsmatrix nach i extrahierten Faktoren (in Diagonale steht 1.0). Andere Symbole siehe oben.

Die Voreinstellung für die Faktorensignifikanz ist "0", d.h. keinerlei Signifikanztest. Bei großer Variablenzahl bedingen die Signifikanztests eine spürbare Verlängerung der Rechenzeit.

Für das Beispiel in Prog30m2 erhalten wir folgende Ergebnisse, wenn wir in das 2. Eingabefeld "1" setzen:

```
Signifikanz der eingegebenen Korrelationsmatrix
(Bartlett-Test der Sphaerizitaet):
-----------------------------------------------
Chi-Quadrat                      106.4719
Freiheitsgrade                    45
Wahrscheinlichkeit (1-p)*100
dass Matrix signifikant           99.9983 %
Determinante der Matrix            0.1485


Signifikanz der extrahierten Faktoren (Test nach Rippe):
--------------------------------------------------------
Chi-Quadrat-Werte je Faktor
   53.5522     44.2504     40.2207

Freiheitsgrade
     45          36           28

Wahrscheinlichkeit (1-p)*100, dass Faktor  signifikant
   81.4275     83.1802     93.3284

Determinate der durch die Faktoren reproduzierten Matrix
    0.3626      0.3105      0.2904
```

Beachte: In die Zahl der signifikanten Faktoren muss der erste nicht mehr signifikante Faktor miteinbezogen werden - sofern ein solcher von Almo noch ausgegeben wird. Ist schon der 1. nicht signifikant, dann ist dieser signifikant (!).
In unserem Beispiel ist also der 1. Faktor - und nur dieser - signifikant.
Die Faktorenzahl, die sich aus dem Kaiser- oder Guttman-Kriterium ergibt und die, die aus dem ststistischen Faktorensignifikanztest hervorgeht, stimmen nicht immer überein.

P30.3.8 Eingabebox: Rotation

Optionsbox geöffnet:

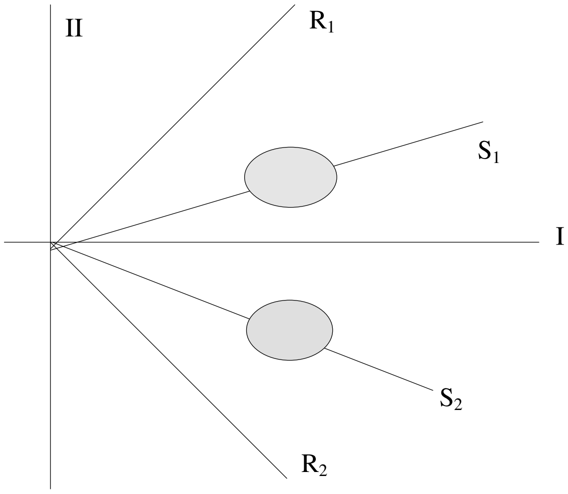
```
[X] Loesche wieder diese Box (dann Voreinstellungen wieder gueltig)

                    Rotation                    Hilfe

                  Zahl der zu rotierenden Faktoren
                  leer = die extrahierten Faktoren werden rotiert
                  0    = nicht rotieren

                  Beispiele für weitere mögliche Angaben:
                  2    =  Rotation für 2 Faktoren
                  2:6  =  Rotation für die ersten 2 Faktoren
                          dann für die ersten 3 Faktoren
                          usw. bis zu den ersten 6 Faktoren

        schiefwinklig   Rotationstyp
                        = schiefwinklig; Quartimin-Rotation mit Ver-
                           besserung durch die Gruppen-Rotation
                           (Voreinstellung)
                        = rechtwinklig;  Varimax-Rotation

     1            Rotationslösung bei schiefwinkliger Rotation
                  1 =die Ergebnisse der durch die "Gruppen-
                     Rotation" verbesserten Quartimin-Rotation
                     werden ausgegeben (Voreinstellung)
                  2 =die reine Quartimin-Lösung wird
                     zusätzlich ausgegeben
                  3 =nur Quartimin-Lösung rechnen und ausgeben

                  4 =Gruppenrotation
                     mit vom Benutzer definierten Variablengruppen
                               ".\TESTDAT\Gruppenzugehoer.mat"
                                          Hilfe
```

Almo führt standardmäßig für die extrahierten Faktoren die schiefwinklige Quartimin-Rotation durch, deren Ergebnisse durch ein sekundäres Rotationsverfahren, die Gruppen-Rotation, verbessert werden. Siehe Abschnitt P30.2.3.2.

Eingabefeld 1: Faktorenzahl
Betrachten wir einige Beispiele:
In das Eingabefeld wird 0 eingegeben. Dann wird nicht rotiert.
Es wird 2 eingegeben. Dann werden nur die ersten 2 Faktoren rotiert, auch wenn mehr als 2 Faktoren extrahiert wurden.
Ein 3. Beispiel: Es wurde 2:4 eingegenben. Dann werden die ersten 2 Faktoren rotiert, dann die ersten 3 Faktoren, dann die ersten 4 Faktoren. Es werden also drei Rotationsösungen ausgegeben. Fer Benutzer kann dann auf Grund inhaltlicher Überlegungen entscheiden, welche Faktorenzahl die angemessene ist.
Im Eingabefeld kann also eine einzelne Zahl, oder 2 durch Doppelpunkt getrennte Zahlen stehen.

Eingabefeld 2: "Rotationstyp"
Möglich sind die beiden Angaben: `schiefwinklig` und `rechtwinklig`.
Die Voreinstellung ist "`schiefwinklig`;" d.h. es wird schiefwinklig rotiert, wenn die Optionsbox "Rotation" geschlossen bleibt.

Der Unterschied zwischen recht- und schiefwinkliger Rotation kann geometrisch gut veranschaulicht werden. Betrachten wir ein Beispiel. Wir zeichnen die Ergebnisse einer Faktorenanalyse mit 2 Faktoren in ein Koordinatensystem



Die Faktorenanalyse liefert 2 deutlich voneinander getrennte Variablengruppen, die wir in der Grafik als 2 „Punktewolken“ durch 2 Ellipsen dargestellt haben.

Die Achsen I und II bilden das rechtwinlige Ausgangs-Koordinatensystem. Die Achsen S_1 und S_2 sind die schiefwinkeligen Achsen, die uns das Quartimin-Verfahren (mit oder ohne anschließender Gruppenrotation) liefert. Sie laufen bei der Quartimin-Rotation ungefähr "mittig", bei der Gruppengenauer "varianzmaximierend" durch die beiden Punktwolken.
Die Achsen R_1 und R_2 sind die rechtwinkeligen Achsen, die uns das rechtwinkelige Varimax-Verfahren liefert. Es ist deutlich zu erkennen, dass je dichter die beiden Punktewolken beieinander liegen, die Varimax- Lösung umso schlechter ist. Wenn keine besonderen Gründe gegeben sind, dann sollte man schiefwinkelig rotieren. Siehe dazu auch die Ausführungen in Abschnitt P30.1.9.

Eingabefeld 3: Rotationslösung bei schiefwinkloger Rotation

1 Die Quartimin-Lösung wird durch die Gruppenrotation verbessert. Nur deren Ergebnisse werden ausgegeben. Dies ist die Voreinstellung.
2 Die Quartimin-Lösung wird zusätzlich ausgegeben.
3 Es wird nur die Quartimin-Lösung ermittelt und ausgegeben.
4 Gruppenrotation mit vom Benutzer definierten Variablengruppen

Eingabefeld 4: Datei der benutzer-definierten Variablengruppen

Die Gruppenrotation kann auch ohne vorausgehende Quartimin-Rotation gerechnet werden. Es genügt wenn der Benutzer die Variablen in so viele Gruppen separiert wie Faktoren rotiert werden sollen. Durch das Gruppenrotations-Verfahren werden dann schiefwinklige "varianzmaximierende" Achsen durch die Variablengruppen (=Punkte-wolken) gezogen.
Im Eingabefeld muss dem Program mitgeteilt werden, in welcher Datei der Benutzer die Gruppenzugehörigkeit der zu faktorisierenden Variablen geschrieben hat.
Wurde nicht die Rotationslösung 4 gewählt, dann ist der im Eingabefeld enthaltene Namen irrelevant. Almo versucht dann nicht die betreffende Datei zu laden.

Betrachten wir ein Beispiel. Siehe dazu Beispielprogramm Prog30m2_Grpz.Alm. Das Programm findet man nach Klick auf den Knopf "alle Progs" im Oberrand des Almo-Fensters.
Im Beispiel werden 6 miteinander korrelierende Variable faktorisiert. Es werden 2 Faktoren extrahiert. Entsprechend sind dann 2 Variablengruppen zu bilden. Der Benutzer weiß ungefähr aus vorausgehenden Faktorenanalysen, wie er die 6 Variablen in Gruppen separieren möchte, beispielsweise so: Gruppe1: V1,2,3,4 Gruppe2: V5,6

Er schreibt folgende Matrix

```
1    0
1    0
1    0
1    0
0    1
0    1
```

Die Matrix hat 6 Zeilen (entsprechend der 6 Variablen) und 2 Spalten (entsprechend der 2 Faktoren/Variablengruppen). Eine 1 bedeutet "Gruppenzugehörigkeit: ja" und 0 "nein". Im Beispiel gehören die ersten 4 Variablen zur Gruppe 1 und die 5. und 6. Variablen zur Gruppe 2. Der Benutzer legt eine neue Datei an, z.B. mit dem Namen "Almo15\Progs\Gruppenzugehoer.mat", in die er diese Matrix speichert.

P30.3.9 Eingabebox: Weitere Optionen (Faktorwert-Koeffizienten, Distanzmatrix)

Optionsbox geöffnet:

Eingabefeld 1:
Faktorwert-Koeffizienten (= Faktor-Betaladungen)
Das Ziel der Faktorenanalyse ist, in aller Regel
(1) die (Mehr-)Dimensionalität einer Variablenmenge zu studieren und
(2) für die einzelnen Dimensionen (= Faktoren) dann Faktorwerte zu berechnen, die den einzelnen Untersuchungseinheiten zugeordnet werden können.

Zur Bildung der Faktorwerte können nicht unmittelbar die Faktorladungen verwendet werden. Diese müssen zuerst auf die sogenannten Faktorwert-Koeffizienten (bzw. Faktor-Betaladungen) transformiert werden.
Wird im Eingabefeld „1“ eingetragen, dann berechnet ALMO Faktorwertkoeffizienten aller extrahierten Faktoren aus der unrotierten Faktorladungsmatrix und zusätzlich aus der rotierten Faktorladungsmatrix (sofern rotiert wurde).

Sätze

1. Im Fall der sogenannten "Hauptkomponenten-Methode", d.h. wenn die Kommunalitäten mit der Voreinstellung geschätzt werden und wenn eine "normale" Faktorenanalyse gerechnet wird, dann werden die Faktorwert-Koeffizienten gemäß folgender Gleichung errechnet:

   $f_{ik} = a_{ik}/E_k$

   dabei ist

   f_{ik}= Faktorwert der Variablen i für Faktor k
   a_{ik}= Faktorladung der Variablen i auf dem Faktor k
   E_k = k-ter Eigenwert

   (siehe dazu Holm, 1976, S.75).

   Auch die Faktorwert-Koeffizienten der nominalen Faktorenanalyse (Abschnitt P30.8) werden nach der hier beschriebenen Methode gebildet.

   Die oben beschriebene Voreinstellung ist aktiv, wenn der Benutzer die Optionsboxen „Kommunalitätenschätzung“ und „Faktorenanalytisches Modell und Verfahren“ nicht öffnet. Oder wenn er in der erst genannten Optionsbox im 1. Eingabefeld (= Kommunalität) „1“ und im 3. Eingabefeld (Kommunalitäten-Iteration) „0“ einsetzt und wenn er in der zweit genannten Optionsbox im 1. Eingabefeld (= faktorenanalytisches Modell) „normale_Fakt“ einträgt.

2. In allen anderen (nicht in 1 beschriebenen Fällen) wird die sogenannte "Regressionsmethode" verwendet. Es gilt folgende Gleichung

   $\mathbf{F} = \mathbf{S} \cdot \mathbf{R}^{-1}$

   dabei ist
   F = Matrix der Faktorwerte-Koeffizienten
   S = Matrix der Faktorladungen. Wird die schiefwinklig rotierte Lösung verwendet, dann ist S die Strukturmatrix (der rechtwinklig auf die schiefwinkligen Achsen projizierten Ladungen)
   $\mathbf{R}^{-1}$=Inverse der Korrelationsmatrix (bzw. der Quadratsummen- oder Kovarianzmatrix). Siehe Harman, 1967, S.351, Holm, 1976, S.76).

3. Die Faktorwert-Koeffizienten sind verschieden, wenn sie aus der Korrelations-, Kovarianz- oder Quadratsummen-Matrix ermittelt werden. Zwischen Kovarianz- und Quadratsummen-Matrix besteht jedoch folgender einfacher Zusammenhang

   $c_i = \sqrt{n} \cdot q_i$

   wobei
   c_i = Faktorwertkoeffizient für Variable i aus Kovarianzmatrix
   q_i = Faktorwertkoeffizient für Variable i aus Quadratsummenmatrix
   n = Zahl der Untersuchungseinheiten.

4. Almo berechnet die Matrix der Faktorwert-Koeffizienten aus der unrotierten Faktorladungsmatrix und (sofern rotiert wird) aus der (rotierten) Faktor-Ladungs- bzw. Strukturmatrix.

Wenn die Faktorwert-Koeffizienten vorliegen, dann können für jede einzelne Untersuchungseinheit deren Faktorwerte in den extrahierten Faktoren ermittelt werden. In der Regel wird man so vorgehen, dass man eine neue Datei bildet, die aus der seitherigen Datei besteht plus den Faktorwert-Variablen.

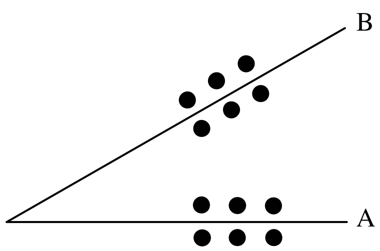
Eingabefeld 2: Distanzmatrix ermitteln
Wird "1" eingegeben, dann werden aus der unrotierten Faktorladungsmatrix die euklidischen Distanzen zwischen den Variablen ermittelt.

Eingabefeld 3: Zwischenergebnisse
Wird "1" eingegeben, dann werden verschiedene Zwischenergebnisse des Kalküls ausgegeben.

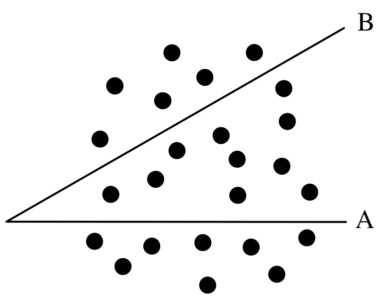
Eingabefeld 4: Eigenständigkeitskoeffizienten
0 =keine
Sollen Eigenständigkeitskoeffizienten berechnet werden, dann empfehlen wir in das Eingabefeld einzusetzen: 90. Das bedeutet, dass der Eigenständigkeitstest mit einer Signifikanz von 90 % gerechnet wird. Zum Begriff der Eigenständigkeit siehe Holm: Befragung 3, Abschnitt 13. Wir wollen hier diesen Begriff kurz erläutern.

Wir wollen annehmen, wir hätten 2 verschiedene Variablenmengen faktorenanalytisch untersucht. Für die eine hätten wir als Ergebnis der „Rotation“ die in Zeichnung 1 und für die andere die in Zeichnung 2 erhaltene Konfiguration erhalten.

Zeichnung 1



Zeichnung 2

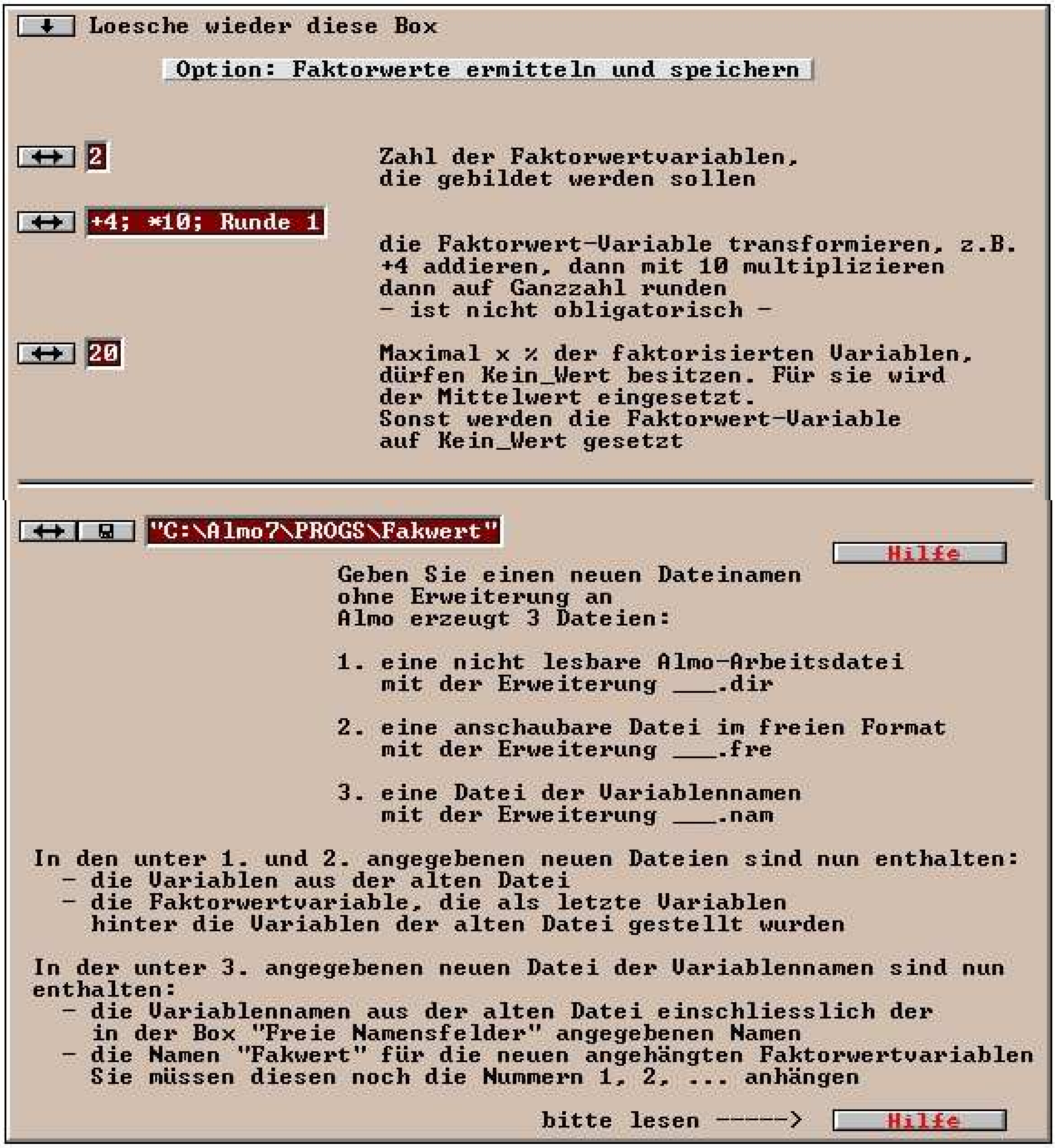


Es ist offensichtlich, dass in Zeichnung 1 zwei eigenständigeVariablengruppen A und B vorhanden sind. In Zeichnung 2 ist zwar auch eine Achse A und B vorhanden (als Ergebnis der Rotation), aber ganz augenscheinlich sind A und B nur rechnerisch entstandene Artefakte. Wir haben eine Punktwolke im zweidimensionalen Raum ohne erkennbare eigenständige Untergruppen. In diesem Falle haben wir die Rotation umsonst durchgeführt. Der 1. Faktor der originalen Faktorladungsmatrix entspricht der Zieldimension. Der 2. Faktor ist eine inhaltlich nicht identifizierte unselbständige Fremddimension.
Wie soll nun entschieden werden, ob eine Punktekonfiguration in eigenständige Untergruppen aufgelöst ist oder nicht.

Der Verfasser hat zu diesem Zwecke einen Eigenständigkeitskoeffizienten entwickelt. Er basiert auf dem bei Überla (1968, S. 184 – 187) beschriebenen Test von Bargman (1955) zur Ermittlung der Signifikanz der Einfachstruktur.

P30.3.10 Eingabebox: Faktorwerte ermitteln und speichern

Optionsbox geöffnet:



Das Speichern der Faktorwert-Variablen hat nur dann einen Sinn, wenn Sie die entgültige Faktoren-Lösung gefunden haben.

Bevor wir die Eingabefelder dieser Eingabebox erläutern, wollen wir den Begriff der Faktorwertvariablen erklären.

Wie wir bereits ausgeführt haben, ist der Zweck der Faktorenanalyse ein doppelter:

1. Es gilt die Dimensionalität (=die Faktorenstruktur) einer Variablenmnge zu studieren
2. Es geht darum, für die einzelnen Faktoren sogenannte Faktorwerte zu ermitteln und diese den einzelnen Untersuchungseinheiten als zusätzliche Variable anzuhängen.

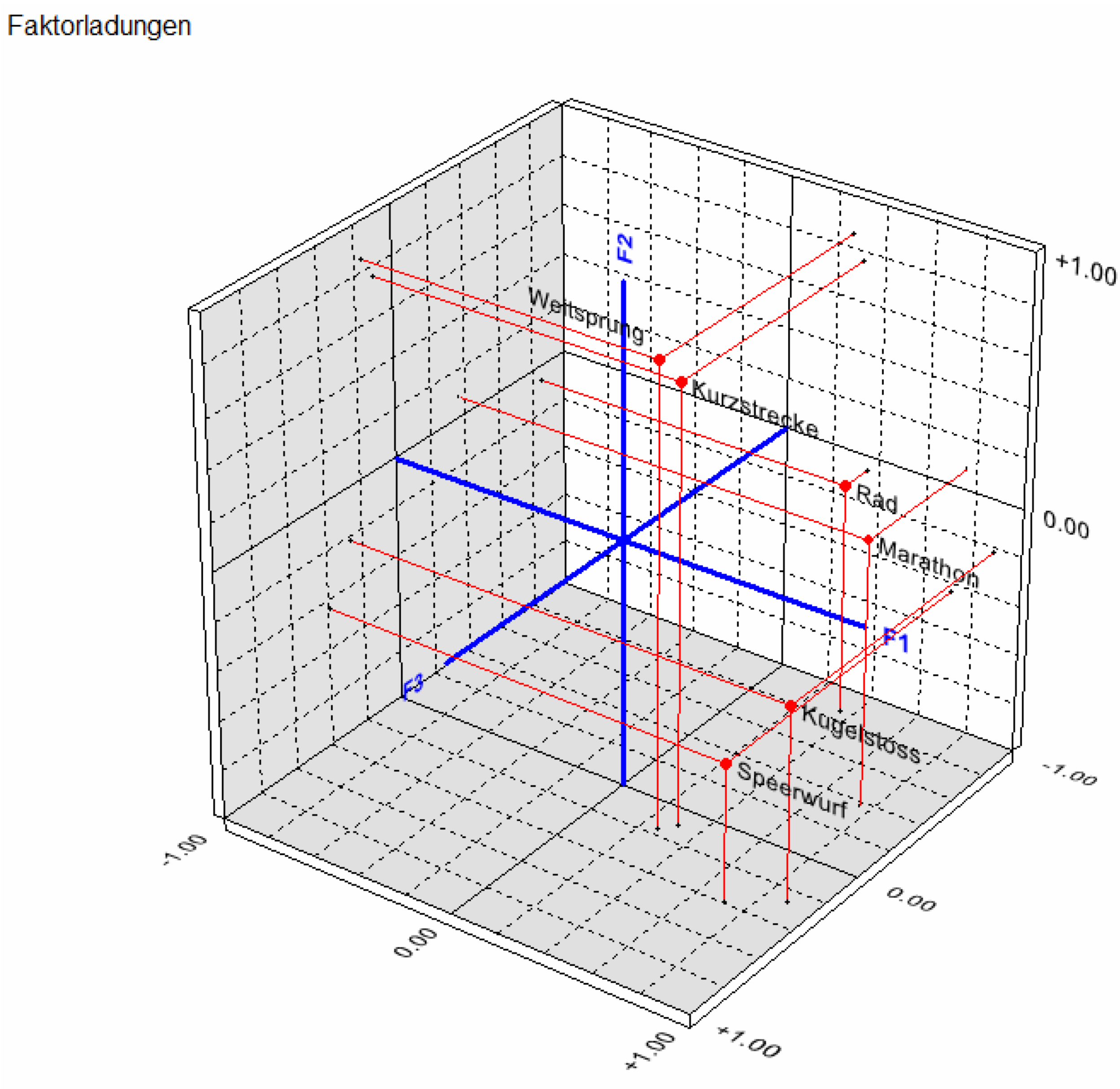
Werden Faktorwerte angefordert, dann ermittelt Almo zuerst die Faktorwert-Koeffizienten (=Faktor-Betaladungen) - auch wenn die oben erläuterte Optionsbox "Weitere Optionen" nicht geöffnet wurde.

In unserem Beispiel liefert Almo folgende Ausgabe, die wir hier stark verkürzt zeigen:

Matrix der Faktorladungen

		Faktor 1	Faktor 2	Faktor 3
Marathon	V1	0.7611	0.0837	-0.3750
Rad	V2	0.3479	-0.0616	-0.8624
Kugelsto	V3	0.8856	-0.2153	0.2636
Speerwur	V4	0.7154	-0.4378	0.3925
Weitspru	V5	0.2642	0.8526	0.1575
Kurzstre	V6	0.3111	0.7597	0.0955

Dies sind die unrotierten Faktorladungen. Almo liefert noch folgende Grafik



Unrotierte Faktor-Betaladungen (Faktorwert-Koeffizienten)(Hauptkomponenten-Loesung)

		Faktor 1	Faktor 2	Faktor 3
Marathon	V1	0.3519	0.0539	-0.3284
Rad	V2	0.1609	-0.0396	-0.7553
Kugelsto	V3	0.4094	-0.1386	0.2308
Speerwur	V4	0.3307	-0.2819	0.3437
Weitspru	V5	0.1221	0.5490	0.1380
Kurzstre	V6	0.1438	0.4892	0.0836

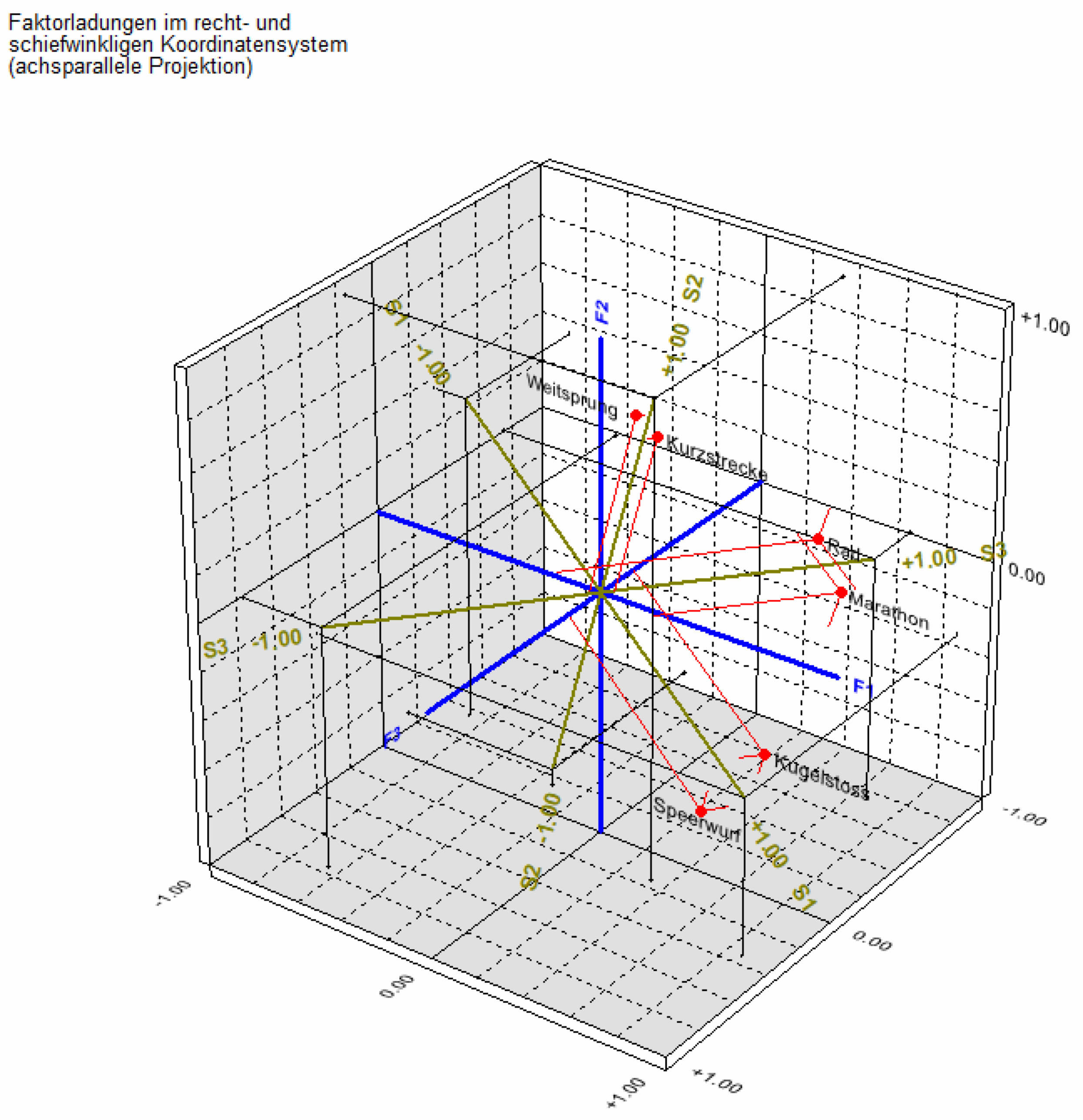
Dies sind die Faktorwert-Koeffizienten die Almo aus den unrotierten Faktorladungen errechnet. Sie werden für die Faktorwertbildung verwendet wenn der Benutzer sich mit einer unrotierten Faktorenlösung zufrieden geben will.

Wird jedoch (schiefwinklig) dann müssen die Faktorwert-Koeffizienten aus den rotierten Faktorladungen ermittelt werden.

Matrix der auf die schiefwinkligen Achsen achsparallel projizierten Faktorladungen (Ladungsmatrix)

		Faktor 1	Faktor 2	Faktor 3
Kugelsto	V3	0.9070	0.1103	0.0969
Speerwur	V4	0.9430	-0.1131	-0.0994
Weitspru	V5	-0.0141	0.9096	-0.0372
Kurzstre	V6	0.0154	0.8204	0.0409
Marathon	V1	0.2995	0.1969	0.6713
Rad	V2	-0.2653	-0.1745	0.9776

Dies sind die rotierten Faktorladungen.
Almo liefert noch folgende Grafik



Die Achsen S1-S2-S3 sind die schiefwinkligen Achsen. Die Punkte sind achsparallel auf die jeweiligen Flächen projiziert, die durch je 2 schiefwinklige Achsen gebildet werden

Offensichtlich ist Faktor S1 die "Kraft", Faktor S2 ist "Sprungkraft" Faktor S3 ist "Ausdauer".

```
Matrix der Winkel zwischen den schiefwinkligen Achsen
```

	Faktor 1	Faktor 2	Faktor 3
Faktor 1	0	89.3727	73.0331
Faktor 2	89.3727	0	83.6725
Faktor 3	73.0331	83.6725	0

Die 3 schiefwinkligen Achsen stehen in relativ stumpfen Winkeln zueinander. D.h. sie repräsentieren 3 relativ unabhängige, eigenständige Messdimensionen, die wir mit "Kraft", "Sprungkraft" und "Ausdauer" inhaltlich identifizieren können.

```
Rotierte Faktor-Betaladungen (Faktorwert-Koeffizienten)
(Hauptkomponenten-Loesung)
```

		Faktor 1	Faktor 2	Faktor 3
Kugelsto	V3	0.4853	0.0402	0.0897
Speerwur	V4	0.5074	-0.1028	-0.0484
Weitspru	V5	-0.0383	0.5725	-0.0210
Kurzstre	V6	-0.0177	0.5159	0.0339
Marathon	V1	0.1698	0.1180	0.4787
Rad	V2	-0.1139	-0.0953	0.6781

Dies sind nun die Faktorwert-Koeffizienten, die Almo aus der obigen "Ladungsmatrix" d. h. den rotierten (und achsparallel projizierten) Faktorladungen errechnet. Sie werden für die eigentliche Faktorwertbildung verwendet.

Wenn der Benutzer die Optionsbox "Weitere Optionen" öffnet, und im 3. Eingabefeld (Zwischenergebnisse ausgeben) "1" einsetzt, dann zeigt Almo ihm, wie die Faktorwerte je Untersuchungseinheit errechnet werden. Almo gibt in diesem Falle zusätzlich aus:

```
Eingelesene Mittelwerte
 105.7800  49.9000  70.5600  47.1800  74.7200  52.0000

Eingelesene Standardabweichungen
 28.2937  30.9453  25.1015  26.4935  28.6273  28.2475
```

Das sind die Mittelwerte und Standardabweichungen der 6 faktorisierten Variablen.

```
Als Beispiel wird die Faktorwert-Berechnung fuer den
1.Datensatz gezeigt

Jede einzelne Variable wird standardisiert (sofern die Korrelations-
Matrix faktorisiert wurde) und mit dem Faktorwert-Koeffizienten
multipliziert. Die Formel ist folgende:
```

```
(Variablenwert - Mittelwert) * FakwKoeff / Standabwg
```

```
Wurde die Kovarianzmatrix faktorisiert, dann ist Standabwg = 1
Wurde die (durchschnittliche) Kreuzproduktematrix faktorisiert,
dann ist Mittelwert = 0 und Standabwg = 1

V1                          (91 - 105.78) * 0.169845 / 28.2937
V2                        + (39 - 49.9) * -0.113873 / 30.9453
V3                        + (48 - 70.56) * 0.485288 / 25.1015
V4                        + (4 - 47.18) * 0.507383 / 26.4935
V5                        + (48 - 74.72) * -0.0383434 / 28.6273
V6                        + (19 - 52) * -0.0177444 / 28.2475
Faktorwert-Variable V7   = -1.2552


V1                          (91 - 105.78) * 0.118046 / 28.2937
V2                        + (39 - 49.9) * -0.0952589 / 30.9453
V3                        + (48 - 70.56) * 0.0401705 / 25.1015
V4                        + (4 - 47.18) * -0.102768 / 26.4935
V5                        + (48 - 74.72) * 0.572548 / 28.6273
V6                        + (19 - 52) * 0.515905 / 28.2475
Faktorwert-Variable V8   = -1.03382


V1                          (91 - 105.78) * 0.478741 / 28.2937
V2                        + (39 - 49.9) * 0.6781 / 30.9453
V3                        + (48 - 70.56) * 0.0896848 / 25.1015
V4                        + (4 - 47.18) * -0.0483817 / 26.4935
V5                        + (48 - 74.72) * -0.0210046 / 28.6273
V6                        + (19 - 52) * 0.0339105 / 28.2475
Faktorwert-Variable V9   = -0.510695
```

Die 3 Faktorwerte werden als V7, V8, V9 an das Ende des Datensatzes angehängt Betrachten wir die 1. Zeile aus der Berechnung der Faktorwervariablen V7.

```
        (91 - 105.78) * 0.169845 / 28.2937

         |      |        |         |
         |      |        |         |
   Variablenwert |     Faktorwert-  Standard-
   von V1       |     Koeffizient  abweichung
          Mittelwert  für V1       von V1
          von V1
```

In entsprechender Weise werden in den nachfolgenden Zeilen, die weiteren faktorisierten Variablen V2 bis V6 behandelt. Die Summe dieser 6 Ausdrücke ergibt dann den Faktorwert V7.

Eingabefeld 1: Geben Sie hier die Zahl der Faktorwervariablen an, die Almo berechnen soll. Diese Zahl sollte mit der Zahl der extrahierten Faktoren Übereinstimmen. Sie muss dies aber nicht. Die Zahl der zu extrahierenden Faktoren können Sie in der Optionsbox "Faktoren" einstellen. In unserem Beispiel sollen 3 "Faktorwert-Variable" gebildet werden, eine für "Kraft", eine Für "Sprungkraft" und eine für "Ausdauer".

Eingabefeld 2: Die Faktorwert-Variablen besitzen Werte von ca. -4 bis +4. Sie können mehrere signifikante Kommastellen haben. Es ist nun durchaus zulässig die Faktorwert-Variablen linear zu transformieren. In unserem Beispiel haben wir +4 addiert (damit keine negativen Werte auftreten), dann mit 10 multipliziert und dann auf Ganzzahl gerundet (damit keine Dezimalwerte auftreten). Das kann man machen, muss es aber nicht. Lesen Sie nochmals unsere obigen Erläuterungen zu den Boxen 7 und 8. Werden Datensätze ein- oder ausgeschlossen oder werden Variable umkodiert, dann wirkt sich das auch auf die Faktorwerte und deren Abspeichern in einer neuen Datei aus.

Eingabefeld 3: Kein-Wert-Behandlung der faktorisierten Variablen. Wenn eine der faktorisierten Variablen, die für die Faktorwert bildung verwendet werden, keinen Wert besitzt - was soll dann geschehen. Almo setzt in diesem Falle den Variablenmittelwert als Variablenwert ein.Das bedeutet im Prinzip, dass diese Variable zum Faktorwert .0 beiträgt.

Der Benutzer kann nun festlegen, wieviele der faktorisierten Variablen Kein-Wert besitzen dürfen. Trägt der Benutzer z.B. "30" ein, dann heißt dies, dass für eine Untersuchungseinheit maximal 30 % der faktorisierten Variablen Kein_Wert besitzen. Das sind 2 von den insgesamt 6 Variablen in unserem Beispiel. Für sie wird der Mittelwert eingesetzt. Sonst (wenn es mehr sind) wird die Faktorwert-Variable für diese Untersuchungseinheit auf Kein_Wert gesetzt.

Setzt der Benutzer "0" ein dann darf keine der faktorisierten Variablen Kein-Wert besitzen. Besitzt auch nur eine Kein-Wert, dann wird die Faktorwertvariable für diese Untersuchungseinheit auf Kein-Wert gesetzt.

Eingabefeld 4: Geben Sie einen Namen für die neue Datei an. In diese werden die seitherigen Variablen geschrieben, sowie die neu gebildeten "Faktorwert-Variablen". Geben Sie den Dateinamen ohne Erweiterung an. Almo erzeugt dann 3 Dateien:

1. eine nicht lesbare Almo-Arbeitsdatei
   mit der Erweiterung ___.dir

2. eine anschaubare Datei im freien Format
   mit der Erweiterung ___.fre

3. eine Datei der Variablennamen
   mit der Erweiterung ___.nam

In den unter 1. und 2. angegebenen neuen Dateien sind nun enthalten:
- die Variablen aus der alten Datei
- die Faktorwertvariable, die als letzte Variablen hinter die Variablen der alten Datei gestellt wurden

In der unter 3. angegebenen neuen Datei der Variablennamen sind nun enthalten:
- die Variablennamen aus der alten Datei einschliesslich der in der Eingabebox "Freie Namensfelder" angegebenen Namen
- die Namen "Fakwert" für die neuen angehängten Faktorwertvariablen (in unserem Beispiel: V7,7,8).

Sie können die Datei der Variablennamen "Fakwert.nam" in ein Fenster laden und die Variablennamen beliebig verändern. Danach wieder speichern.

In unserem Beispiel enthält die Datei "Fakwert.nam" folgende Namen:

```
Name 1=Marathon;
Name 2=Rad;
Name 3=Kugelstoss;
Name 4=Speerwurf;
Name 5=Weitsprung;
Name 6=Kurzstrecke;
```

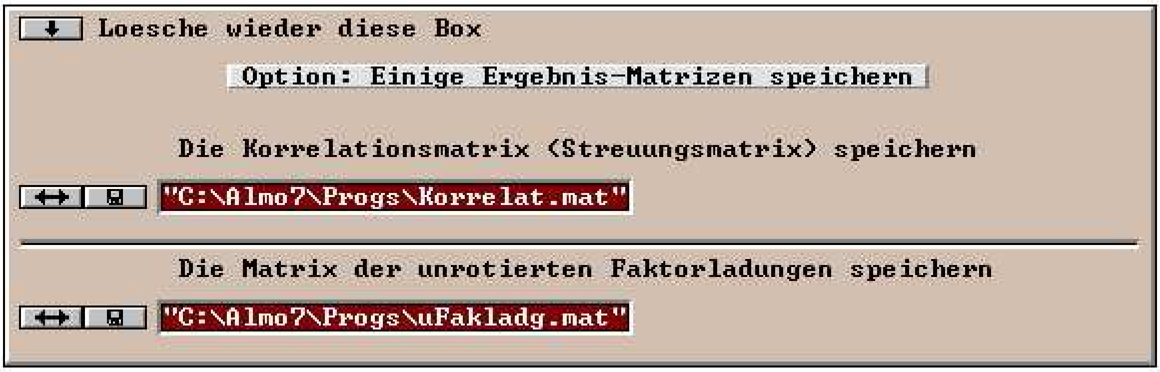
```
Name 7=Fakwert;
Name 8=Fakwert;
Name 9=Fakwert;
```
Almo verwendet teilweise eine Kurzschreibweise. Es schreibt z.B. anstelle von "Name 1=..." die zulässiger Kurzform "N1=....".

Die 3 neuen Faktorwertvariablen V7,8,9 heißen alle "Fakwert". Sie müssen diese umbenennen damit sie sich unterscheiden. Im einfachsten Fall hängen Sie die Nummern 1, 2, 3 an oder die Nummer der Faktorwertvariablen, also 7, 8, 9. Sinnvoll ist es jedoch den Faktorwertvariablen, einen Namen zu geben, der ausdrückt, was sie messen - beispielsweise so:

```
Name 7=Kraft;
Name 8=Sprungkraft;
Name 9=Ausdauer;
```

P30.3.11 Eingabebox: Einige Ergebnis-Matrizen speichern

Optionsbox geöffnet:



Eingabefeld 1: Almo errechnet zuerst eine Streuungsmatrix, auf die es dann den Kalkül der Faktorenanalyse anwendet. In der Regel ist dies die Korrelationsmatrix. Der Benutzer kann in der Eingabebox 13 "Zu faktorisierende Matrix" aber auch andere Streuungsmatrizen einstellen. Die Korrelationsmatrix kann als "Ähnlichkeitsmatrix" begriffen werden und beispielsweise einer nichtmetrischen multidimensionalen Skalierung nach Kruskal unterworfen werden. Siehe dazu Programm 34 bzw. unter "Verfahren / MDS: Nichtmetrische multidimensionale Skalierung". Wird das Eingabefeld leer gemacht, dann wird nicht gespeichert.

Eingabefeld 2: Die unrotierte Faktorladungsmatrix kann gespeichert werden. Wird das Eingabefeld leer gemacht, dann wird nicht gespeichert.

P30.3.12 Eingabebox: "Aussehen" der auszugebenden Tabelle bzw. Matrix
Siehe dazu P0.9

P30.3.13 Eingabebox 21: Grafik-Optionen
Siehe dazu P0.10

P30.4 Faktorenanalyse mit eingegebener Matrix

P30.4.1 Programm-Masken für eingegebene Matrix

Die zu faktorisierende Matrix kann auch direkt eingegeben werden, d.h vom Benutzer selbstgeschrieben werden.

Wir werden im Folgenden ein Maskenprogramm ohne und eines mit Optionen wiedergeben. Sie finden diese Programme durch Klick auf „Verfahren/Faktorenanalyse/Prog30m3 bzw. Prog30m4“.

Prog30m3.Msk
Faktorenanalyse
mit eingegebener fertiger Korrelationsmatrix

Beispiel: Es soll festgestellt werden, ob 6 verschiedene sportliche Leistungen auf einer bzw. mehreren Fähigkeiten (=Faktoren) beruhen.

Almo setzt als Kommunalitätenschätzung die multiplen Bestimmtheitsmaße (=R-Quadrat) ein. Dann wird zuerst die unrotierte Faktorladungsmatrix ermittelt.
Diese kann als 2- bzw. 3-dimensionales Koordinatensystem grafisch dargestellt werden. Dann wird schiefwinklig rotiert. D.h. es wird versucht, räumlich voneinander getrennte "Punktewolken" zu identifizieren und durch diese Achsen zu ziehen. Diese Achsen können dann als die geometrischen Repräsentanten der (in unserem Beispiel) Fähigkeiten betrachtet werden, die hinter den 6 sportlichen Leistungen stehen.

Siehe Handbuch, Abschnitt P30

Was ist ein Kurzprogramm ? --> Hilfe
Maskenbedienung --> Hilfe

1 Speicher fuer x Variable
Geben Sie soviele an, wie die zu faktorisierende Matrix Zeilen bzw. Spalten besitzt

Vereinbare Variable= 6 ;

2 Option: Weitere Vereinbarungen - nur wenn Almo dazu auffordert

3 Den Variablen Namen geben

Name1=Marathon;
Name2=Rad;
Name3=Kugelstoss;
Name4=Speerwurf;
Name5=Weitsprung;
Name6=Kurzstrecke;

erzeuge zusätzliche Namensfelder

4 Zahl der Zeilen bzw. Spalten der Matrix

6

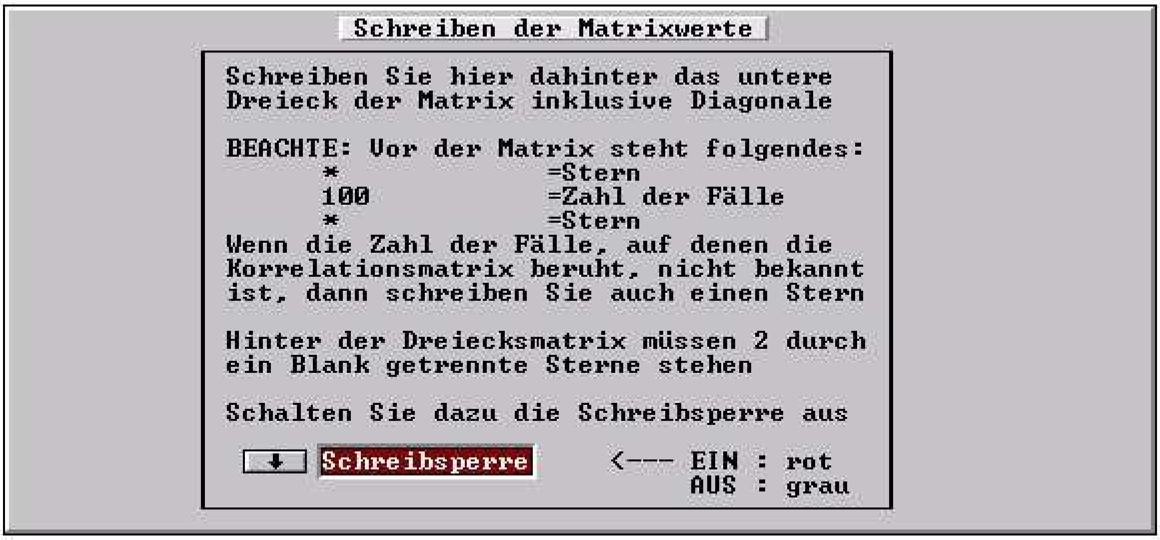
5 Matrix aus Datei oder "selbst geschrieben" Hilfe

Eingabe

1. Möglichkeit: Benutzer will Matrix selbst schreiben
Klick auf den 3. Knopf. Almo setzt das Wort "Eingabe" ein
Schalten Sie am Programmende die Schreibsperre aus
und schreiben Sie die Matrix-Daten hinter dem Programm

2. Möglichkeit: Matrix soll aus Datei eingelesen werden
Klicken Sie auf den 2. Knopf und selektieren die Datei
Schalten Sie in der nächsten Box die Schreibsperre aus
und löschen Sie die Daten hinter dem Programm

6




```
*
100
*

1.0
0.4   1.0
0.5   0.02  1.0
0.2  -0.07  0.7   1.0
0.3  -0.1   0.1  -0.1   1.0
0.07  0.04  0.1  -0.05  0.6  1.0
*
*
```

Prog30m4.Msk
Faktorenanalyse
mit eingegebener fertiger Matrix
mit Optionen

Beispiel: Es soll festgestellt werden, ob 6 verschiedene sportliche Leistungen auf einer bzw. mehreren Fähigkeiten (=Faktoren) beruhen.

Eine Standard-Faktorenanalyse würde so verlaufen:

Almo ermittelt die unrotierte Faktorladungsmatrix. Diese kann als 2- bzw. 3-dimensionales Koordinatensystem grafisch dargestellt werden. Dann wird schiefwinklig rotiert. D.h. es wird versucht, räumlich voneinander getrennte "Punktewolken" zu identifizieren und durch diese Achsen zu ziehen. Diese Achsen können dann als die geometrischen Repräsentanten der (in unserem Beispiel) Fähigkeiten betrachtet werden, die hinter den 6 sportlichen Leistungen stehen.

Über eine Vielzahl von Optionen, kann diese Standard-Vorgehensweise modifiziert werden. Kurz und schlagwortartig:

Faktorenanalyse als Hauptachsen-Lösung, Hauptkomponenten-Lösung, Alpha-Faktorenanalyse, Image-Faktorenanalyse, kanonische Faktorenanalyse.

Faktorenanalyse angewandt auf die Korrelations- oder Kovarianzmatrix oder auf die Matrix der Abweichungsquadrate

Rechtwinklige Varimax-Rotation oder schiefwinklige Quartimin-Rotation.

Kommunalitätenschätzung: Multiple Bestimmtheitsmasse, oder 1.0 oder größte Korrelation oder selbst gewählte Werte. Iteration der Kommunalitäten.

Faktoren-Signifikanz kann geprüft werden. Festlegen der Faktorenzahl. Ermitteln von Faktorwert-Koeffizienten. Berechnen der euklidischen Distanzen zwischen den Variablen

Siehe Handbuch, Abschnitt P30

Was ist ein Kurzprogramm ? --> Hilfe
Bedienung --> Hilfe

1 Speicher fuer x Variable
Geben Sie soviele an, wie die zu faktorisierende Matrix Zeilen bzw. Spalten besitzt

Vereinbare Variable= 6 ;

2 Option: Weitere Vereinbarungen - nur wenn Almo dazu auffordert

3 Den Variablen Namen geben

Name1=Marathon;
Name2=Rad;
Name3=Kugelstoss;
Name4=Speerwurf;
Name5=Weitsprung;
Name6=Kurzstrecke;

erzeuge zusätzliche Namensfelder

4 Zahl der Zeilen bzw. Spalten der Matrix

6

5 Matrix aus Datei oder "selbst geschrieben" Hilfe

Eingabe

1. Möglichkeit: Benutzer will Matrix selbst schreiben
Klick auf den 3. Knopf. Almo setzt das Wort "Eingabe" ein
Schalten Sie am Programmende die Schreibsperre aus
und schreiben Sie die Matrix-Daten hinter dem Programm

2. Möglichkeit: Matrix soll aus Datei eingelesen werden
Klicken Sie auf den 2. Knopf und selektieren die Datei
Schalten Sie in der nächsten Box die Schreibsperre aus
und löschen Sie die Daten hinter dem Programm

6 Typ der einzulesenden Matrix

Korrelation

=Korrelation Korrelationsmatrix
=Quadratsumme Quadratsummenmatrix
=Kovarianz Kovarianzmatrix
=Quasikorrelation s. Handbuch, P20.8.5.1

7 Option: Faktorenanalytisches Modell und Verfahren

8 Option: Kommunalitätenschätzung

9 Option: Faktoren

10 Option: Rotation

11 Option: weitere Optionen

12 Option: Einige Ergebnis-Matrizen speichern

13 Option: "Aussehen" der auszugebenden Tabelle bzw. Matrix

14 Grafik-Optionen

```
Schreiben der Matrixwerte

Schreiben Sie hier dahinter das untere
Dreieck der Matrix inklusive Diagonale

BEACHTE: Vor der Matrix steht folgendes:
      *              =Stern
      100            =Zahl der Fälle
      *              =Stern
Wenn die Zahl der Fälle, auf denen die
Korrelationsmatrix beruht nicht bekannt
ist, dann schreiben Sie auch einen Stern

Hinter der Dreiecksmatrix müssen 2 durch
ein Blank getrennte Sterne stehen

Schalten Sie dazu die Schreibsperre aus

[↓] Schreibsperre        <--- EIN : rot
                              AUS : grau
```

```
*
100
*

1.0
0.4   1.0
0.5   0.02  1.0
0.2  -0.07  0.7   1.0
0.3  -0.1   0.1  -0.1   1.0
0.07  0.04  0.1  -0.05  0.6   1.0
*
*
```

P30.4.1.1 Erläuterungen zu Prog30m3

Eingabebox 1: Speicher für x Variable

```
Speicher fuer x Variable
Geben Sie soviele an, wie die zu faktorisierende Matrix
Zeilen bzw. Spalten besitzt

Vereinbare Variable= 6 ;
```

Vereinbaren Sie soviele Variable wie die zu faktorisierende Matrix Zeilen bzw. Spalten besitzt.

Eingabebox 2: Weitere Vereinbarungen
Siehe dazu Almo-Dokument Nr. 0 "Arbeiten mit Almo", Abschnitt P0.2.

Eingabebox 3: Den Variablen Namen geben

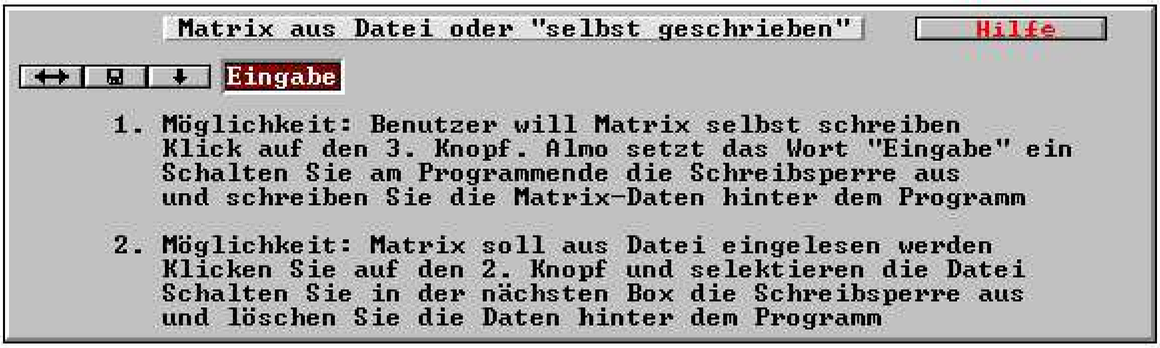
```
Den Variablen Namen geben

[↔] Name1=Marathon;
[↔] Name2=Rad;
[↔] Name3=Kugelstoss;
[↔] Name4=Speerwurf;
[↔] Name5=Weitsprung;
[↔] Name6=Kurzstrecke;

[...] erzeuge zusätzliche Namensfelder
```

Geben Sie den Variablen, die die Matrix bilden, Namen. Nummerieren Sie die Namen durch beginnend mit 1, 2, 3, ... Siehe auch P0.3

Eingabebox 4: Zahl der Zeilen bzw. Spalten der Matrix

Sie müssen hier die Zahl der Zeilen bzw. Spalten der Matrix eintragen.

Eingabebox 5: Matrix aus Datei oder "selbst geschrieben"



Das Programm erlaubt es dem Benutzer, die Matrix am Programmende selbst zu schreiben
Oder die Matrix aus einer Datei einzulesen.

1. Möglichkeit: Benutzer will die Matrix selbst schreiben. Klicken Sie auf den 3. Knopf. Es wird dann in das Eingabefeld das Wort "Eingabe" eingesetzt. Almo erwartet dann die Daten der Matrix am Programmende. Schalten Sie am Programmende in der Box "Schreiben der Matrixwerte" die Schreibsperre aus und schreiben Sie die Matrix-Daten hinter dem Programm.

   Die Matrix-Daten müssen in folgender Form geschrieben werden. Siehe dazu auch Handbuch, Teil 2, Abschnitt 43

6	Grösse der Matrix
487	Zahl der Fälle, aus der die Matrix, z.B. als Korrelationsmatrix gebildet wurde
1.01 1.02 2 3 4 12	Variablennummern. V1 ist nominal mit den beiden Dummies 1.01 und 1.02
1.0	
0.4 1.0	
0.3 0.7 1.0	untere Dreiecksmatrix
0.1 0.2 0.7 1.0	(mit Diagonale)
0.2 0.3 0.2 0.3 1.0	
0.3 0.2 0.2 0.1 0.8 1.0	
0.7 0.6 4.0 1.7 2.5 2.1	Mittelwerte
0.4 0.3 1.2 0.9 0.5 0.9	Stand.abweichungen

   BEACHTE: Jede Zeile, ausser der Dreiecksmatrix, kann durch '*', d.h. einen Stern bzw. das Multiplikationszeichen ersetzt werden. Die Matrix kann also auch so geschrieben werden:

```
*
*
*
1.0
```

0.4 1.0
0.3 0.7 1.0
0.1 0.2 0.7 1.0
0.2 0.3 0.2 0.3 1.0
0.3 0.2 0.2 0.1 0.8 1.0
*
*

2. Möglichkeit: Die Matrix soll aus Datei eingelesen werden. Klicken Sie auf den 2. Knopf. Almo präsentiert die Datei-Auswahl-Box. Selektieren Sie die Datei, die die Matrix enthält. Die Matrix muss die oben dargestellte Form besitzen !! Das ist automatisch der Fall, wenn sie mit einem Almo-Programm gespeichert wurde, beispielsweise mit dem Korrelationsprogramm Prog19bm (Optionsbox: Schreibe errechnete Matrix in Datei). Schalten Sie am Programmende in der Box "Schreiben der Matrixwerte" die Schreibsperre aus und löschen Sie die Daten hinter dem Programm. Mit der Tastenkombination Strg+Y löschen Sie jeweils die ganze Zeile, in der sich der Cursor befindet. Wenn Sie vergessen zu löschen, dann rechnet Almo trotzdem richtig, bringt jedoch zum Ende der Ergebnis- ausgabe die Fehlermeldung, dass mit der Programmstruktur etwas nicht stimmen könnte. Diesen Hinweis können Sie negieren.

Eingabebox 6: Schreiben der Matrixwerte
Siehe dazu oben 1. Möglichkeit und nachfolgend die Erläuterung zu Eingabebox 15 bei Prog30m4.

P30.4.1.2 Erläuterungen zu Prog30m4

Die **Eingabeboxen 1** bis **5** sind dieselben wie beim oben erläuterten Prog30m3.

Eingabebox 6: Typ der einzulesenden Matrix

Siehe hierzu auch P30.3.3 in den Erläuterungen zu Prog30m2. Sie müssen den Typ der Matrix, die Sie eingeben wollen, eingeben. Möglich sind folgende Eingaben

```
Eingabe           Matrix, die Sie schreiben
----------------  -------------------------
Korrelation       =Korrelationsmatrix
Quadratsumme      =Quadratsummenmatrix
Kovarianz         =Kovarianzmatrix
Quasikorrelation  =siehe hierzu Handbuch, P45.12.4.2 und P20.8.5.1
```

Eingabebox 7: Option: Faktorenanalytisches Modell und Verfahren
Siehe P30.3.2 in den Erläuterungen zu Prog30m2.

Eingabebox 8: Option: Kommunalitätenschätzung
Siehe P30.3.4 in den Erläuterungen zu Prog30m2.

Eingabebox 9: Option: Faktoren
Siehe P30.3.5 in den Erläuterungen zu Prog30m2.

Eingabebox 10: Option: Rotation
Siehe P30.3.6 in den Erläuterungen zu Prog30m2.

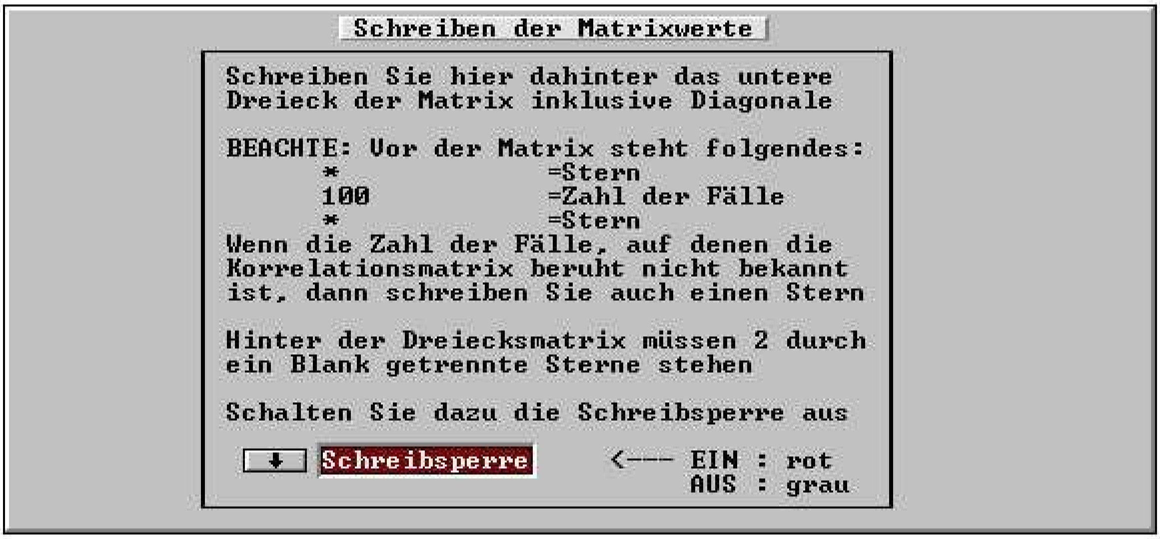
Eingabebox 11: Option: Weitere Optionen
Siehe P30.3.7 in den Erläuterungen zu Prog30m2.

Eingabebox 12: Option: Einige Ergebnis-Matrizen speichern
Siehe P30.3.9 in den Erläuterungen zu Prog30m2.
Beachte: Sie haben hier die Möglichkeit, die Matrix, die Sie mühsam geschrieben haben, in eine dauerhafte Datei zu speichern. Tragen Sie im 1. Eingabefeld dieser Eingabebox, einen Dateinamen ein.

Eingabebox 13: "Aussehen" der auszugebenden Tabelle bzw. Matrix
Siehe dazu Almo-Dokument Nr. 0 "Arbeiten mit Almo", Abschnitt P0.9.

Eingabebox 14: Grafik-Optionen
Siehe dazu Almo-Dokument Nr. 0 "Arbeiten mit Almo", Abschnitt P0.10.

Eingabebox 15: Schreiben der Matrixwerte



Schalten Sie zuerst die Schreibsperre aus, in dem Sie auf den Knopf mit dem nach unten weisenden Pfeil klicken. Hinter der Box werden dan die Daten der Matrix geschrieben.

Die Matrix muss in folgender Form geschrieben werden. Siehe hierzu auch Handbuch, Teil 2, Abschnitt 43.

```
*              ein Stern bzw. Multiplikationszeichen

487            Zahl der Fälle, aus der die Matrix, z.B.
               als Korrelationsmatrix gebildet wurde

*              ein Stern bzw. Multiplikationszeichen

1.0
0.4 1.0
0.3 0.7 1.0                   untere Dreiecksmatrix
0.1 0.2 0.7 1.0               mit Diagonale
0.2 0.3 0.2 0.3 1.0
0.3 0.2 0.2 0.1 0.8 1.0

*              ein Stern bzw. Multiplikationszeichen
*              ein Stern bzw. Multiplikationszeichen
```

Die Zahl der Fälle kann ebenfalls durch einen Stern ersetzt werden. Almo kann dann allerdings keinen statistischen Faktoren-Signifikanztest rechnen, auf den man durchaus verzichten kann. Siehe hierzu P30.3.5.2

Die Matrix kann also legal auch so geschrieben werden:

```
*
*
*
1.0
0.4 1.0
0.3 0.7 1.0
0.1 0.2 0.7 1.0
0.2 0.3 0.2 0.3 1.0
0.3 0.2 0.2 0.1 0.8 1.0
*
*
```

Anhang 1: Kanonische Faktorenanalyse, Alpha-Faktorenanlyse, Image-Faktorenanalyse

Der in diesem Anhang enthaltene Text ist aus meinem Buch "Kurt Holm: Die Befragung 3, Faktorenanalyse", Seite 78-83,85-93 entnommen. An vielen Stellen wurden Anpassungen und kleinere und größere Erweiterungen vorgenommen.

A1.1 Die kanonische Faktorenanalyse

H. Hotelling hat in den 30er Jahren die Methode der kanonischen Korrelation entwickelt. Im Almo-Dokument Nr. 4 "Kanonische Analysen" haben wir dieses Verfahren ausführlich dargestellt. Das Prinzip der kanonischen Korrelation sei hier nochmals kur verdeutlicht:
Bei der kanonischen Korrelation werden 2 Variablengruppen miteinander korreliert:

x_1, x_2, x_3 seien die Variablen der 1. Variablengruppe
y_1, y_2, y_3 seien die Variablen der 2. Variablengruppen

Die Zahl der Variablen in beiden Gruppen darf verschieden sein.
Es werden nun folgende Linearkombinationen gebildet:

(1) $X = \alpha_1 . x_1 + \alpha_2 . x_2 + \alpha_3 . x_3$
(1a) $Y = \beta_1 . y_1 + \beta_2 . y_2 + \beta_3 . y_3$

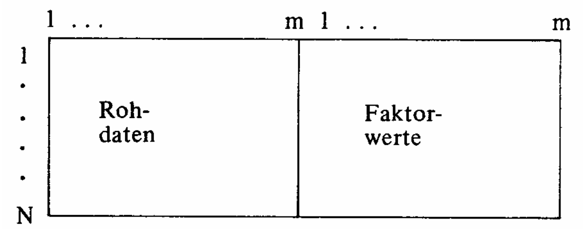
Die Koeffizienten α_1 , α_2 , α_3 bzw. β_1 , β_2 , β_3 nennen wir kanonische Gewichtszahlen. Die beiden gewichteten Summen X und Y nennen wir kanonische *Faktorwertvariable*. So wird für jeden der N Probanden ein *Faktorwer*t X und Y ermittelt. Das Spezifikum der kanonischen Analyse besteht nun darin, dass die Gewichtzahlen α_i und β_j so gewählt werden, daß die beiden Faktorwertvariablen X und Y maximal miteinander korrelieren.

Nun besteht die Möglichkeit einen 2. Satz von kanonischen Gewichtszahlen zu bestimmen, der zum 1. orthogonal ist. So läßt sich dann für alle N Probanden ein 2. Faktorwert X_{II} und Y_{II} errechnen. Diese 2. kanonischen Faktorwertvariablen korrelieren miteinander maximal - mit den kanonischen Faktorwertvariablen X_I und Y_I der 1. Lösung jedoch mit 0. Maximal lassen sich bei m Variablen (in der kleineren Gruppe) m Faktorwertvariable X_1 bis X_m und Y_1 bis Y_m ermitteln.

Rao (1955) hat dieses Verfahren für die Faktorenanalyse aufgegriffen. Seine Methode geht von zweierlei Daten aus:

1) Die Rohdaten, die bei der Untersuchung gewonnen wurden. Sie lassen sich als eine N * m-Matrix darstellen, wobei N = Probandenzahl und m = Zahl der Variablen.
2) Diese Variablen messen auf bestimmten Faktoren. Jede Variable mißt in einem bestimmten Maße auf einem 1., 2., 3., . .bis maximal m-ten. Faktor. D.h. jede der m Variablen miß in bestimmtem unterschiedlichen Maße auf den m Faktoren. Wir unterstellen, wir würden die Werte der N Probanden auf diesen m voneinander unabhängigen Faktoren kennen. Wir besäßen dann eine N * m Matrix der *Faktorenwerte.*

Die beiden Matrizen könnten nun „nebeneinander gestellt“ werden.



Wir wollen als Beispiel drei Rohwert~Variable und dementsprechend drei Faktorwert-Variable unterstellen. Den Rohwertvariablen können die Gewichtszahlen al , a2, a3 und den Faktorwertvariablen die Gewichtszahlen bl , b2 , b3 zugeordnet werden.

Die Rohwerte aus den drei Variablen des 1. Probanden werden gewichtet addiert. Es entsteht der gewichtete Gesamtpunktwert A1 für den l. Probanden. Dann werden die Rohwerte des 2. Probanden gewichtet addiert. Dabei entsteht A2 der gewichtete Gesamtpunktwert der 2. Probanden .... usw.
Entsprechendes geschieht nun auch auf der Seite der Faktorwerte, von denen wir annehmen, dass wir sie kennen. W1 ist der gewichtete Gesamtpunktwert des 1. Probanden, der aus den 3. Faktorwert-Variablen entstand ..... usw. Es ist nun möglich, die gewichteten Gesamtpunktwerte A und W über alle Probanden zu korrelieren.
Die Frage, die nun die kanonische Faktorenanalyse charakterisiert, lautet:

> Wie müssen die Gewichtszahlen a und b gewählt werden, damit A und W maximal miteinander korrelieren?

Wie kann dann ein zweiter Vektor von jeweils a- und b-Gewichten gewählt werden, der zum ersten orthogonal (unabhängig) ist, mit der Bedingung, dass dann ein maximaler Korrelationskoeffizient für A und W entsteht? Wie kann ein dritter Vektor von a- und b-Gewichten gewählt werden, der zu den beiden vorausgehenden orthogonal ist, mit der Bedingung, dass für A und W eine maximale Korrelation entsteht .... usw.

Die kanonische Faktorenanalyse ergibt also Faktorwertvariable, die von den beobachteten Rohwerten mit maximaler Präzision prognostiziert werden. Bei der "normalen" Faktorenanalyse, die sich in ähnlicher Weise entwickeln ließe. werden die b-Gewichte so gewählt, dass die gewichteten Faktorwerte ein Maximum an Varianz in den Rohwerten X erklären.

Die Ableitung der kanonischen Faktorenanalyse verläuft nun in folgender Weise: Die beiden Datensätze. die Rohwertvariablen und die Faktorwertvariablen. wie wir sie oben schematisch dargestellt haben, werden (ohne Gewichtszahlen) interkorreliert.

Wir unterstellen. dass die Rohwert- und Faktorwert-Matrizen standardisierte Daten beinhalten. so dass Kovarianz- und Korrelationsmatrix identisch sind. Diese Korrelationsmatrix besteht aus 4 Submatrizen.

R	**F**
F'	**I**

R ist die m * m Korrelationsmatrix der Rohdaten.

I ist die m * m Korrelationsmatrix der Faktorenwerte. Da die m Faktorwertvariablen voneinander unabhängig sind, ist **I** die Einheitsmatrix (mit 1.0 in der Diagonalen, alle anderen Glieder sind .0)

F ist die m * m Korrelationsmatrix zwischen den Faktorwerten und den Rohdaten.

F' ist die transponierte Matrix zu **F**.

Betrachten wir die l. Spalte von **F**. Sie enthält die Korrelationen der m Variablen aus den Rohdaten mit den Faktorwerten des 1. Faktors. Faktorenanalytisch gesprochen heißt das: In der 1. Spalte von **F** stehen die Faktorladungen der m Variablen auf dem 1. Faktor. **F** ist also die m*m-Faktorladungsmatrix.

Die Matrixmultiplikation **F*F'** entspricht der Multiplikation der Faktorladungsmatrix mit ihrer Transponierten, wobei die Korrelationsmatrix reproduziert wird. Wir können also die Korrelationsmatrix **Rh** mit Kommunalitätenschätzungen in der Diagonalen als eine Annäherung an **F * F'** betrachten.

Die Korrelationsmatrix **Rh** können wir auch durch $\mathbf{R} - \mathbf{U}^2$ bezeichnen. Wir ziehen von der Matrix **R** die Diagonalmatrix der Einzelreste $\mathbf{U}^2$ ab. So entsteht **Rh** . Wir können nun schreiben:

$$(2)\quad \mathbf{R_h} = \mathbf{R} - \mathbf{U}^2 = \mathbf{F} \cdot \mathbf{F}'$$

$\mathbf{U}^2$ ist eine Diagonalmatrix. In ihrer Diagonalen stehen die Einzelreste

$$u_{ii}^2 = 1.0 - h_{ii}^2$$

Almo verwendet standardmäßig als Kommunalität h^2 das multiple Bestimmtheitsmaß.

$\mathbf{U}^{-1}$ ist eine Diagonalmatrix. In ihrer Diagonalen ist der Kehrwert der Wurzel der Einzelreste u_i^2 enthalten.

Nach mehrfachen Umformungen, die wir hier nicht darstellen und begründen, gelangt man zu der Determinanten-Gleichung:

$$(3)\quad | U^{-1} \; (R - U^2) \; U^{-1} - \lambda^* \, I \, | \, c = 0$$

Es gilt also die Eigenwerte λ_j* und die jeweils dazugehörenden normierten Eigenvektoren c_j der Matrix

$$(3a)\quad U^{-1} \; (R - U^2) \; U^{-1}$$

zu bestimmen.

Das geschieht in unserem Computer-Programm mit Hilfe des Tridiagonal-QR-Algorithmus (voreingestellt) oder dem v. Mises-Verfahren. Die Eigenwerte λ_j* sind nicht wie bei der „gewöhnlichen" Faktorenanalyse identisch mit der „Varianz je Faktor".

Wir geben der in 3a definierten Matrix den Namen (in Ermanglung eines besseren) "Kanon.Fakt.Matrix", den wir in der Ausgabe von Zwischenergebnissen verwenden. Die Matrix ist, wie die Korrelationsmatrix eine symmetrische m*m-Matrix (m=Zahl der Variablen).

Wir bezeichnen die m*p Matrix der p normierten Eigenvektoren mit **C** und die p*p Diagonalmatrix der Eigenwerte mit **D**.
Der Vektor der Faktorladungen eines Faktors j ergibt sich aus

$$(4)\quad f_j^* = c_j \sqrt{\lambda_j^*}$$

Aus diesen Vektoren der Faktorladungen wird die m * p Faktorladungsmatrix **F*** gebildet.
Wir können diese auch als ein Matrixprodukt darstellen

$$(5)\quad F^* = C \cdot D^{\frac{1}{2}}$$

In **F*** sind „uneigentliche" Faktorladungen enthalten. Die Korrelationsmatrix $\mathbf{R_h}$ wurde ja mit $\mathbf{U^{-1}}$ prä- und postmultipliziert. Das wird nun wieder „zurückgenommen".

$$(6)\quad F = U \cdot F^* = U \cdot C \cdot D^{\frac{1}{2}}$$

F ist nun die Matrix der „eigentlichen" Faktorladungen. Die quadrierte kanonische Korrelation k eines Faktors j hinsichtlich der m Variablen ist

$$(7)\quad k_j = \frac{\lambda_j^*}{\lambda_j^* + 1}$$

Damit die kanonische Korrelation eines Faktors hinsichtlich aller Variablen mindestens gleich 0 wird, muß λ_j* mindestens gleich .0 sein.
Das ergibt ein wichtiges Kriterium für die praktische Berechnung der Faktorenzahl. Hier wird nun auch die starke formale Ähnlichkeit zur Alpha-Faktorenanalyse (siehe nächsten Abschnitt) sichtbar. Dort geht es darum, den Alpha-Koeffizienten für einen Faktor zu maximieren und nur solche Faktoren zu extrahieren, die einen Alpha-Wert größer 0 haben.
Hier geht es darum, die kanonische Korrelation k eines Faktors hinsichtlich der Variablen zu maximieren, und nur jene Faktoren zu extrahieren, deren k größer 0 ist. In unserem Computer-Programm wird so verfahren. Der Benutzer kann diese Voreinstellung von Almo durchbrechen. Er muss dazu die Optionsbox " Kommunalitätenschätzung" öffnen. Siehe dazu ausführlich P30.3.4.1.

Als Test für die statistische Signifikanz der extrahierten Faktoren wird der Lawley-Test verwendet.

Der Rechengang der kanonischen Faktorenanalyse

Wir werden im Folgenden anhand eines Beipiel die Rechenschritte durch Zahlen veranschaulichen

Wir verwenden die Daten der 6 Sportarten aus Abschnitt P30.2.3.4. Wird in Prog30m2.Msk die Optionsbox "weitere Optionen" geöffnet und "Zwischergebnisse" angefordert, dann gibt Almo diese Zahlen aus. Die Ausgangs-Korrelationsmatrix ist

ursprüngliche Korrelationmatrix R

	Marathon	Rad	Kugelsto	Speerwu	Weitspr	Kurzstr
Marathon	1.0000	0.3975	0.5281	0.2535	0.2727	0.0827
Rad	0.3975	1.0000	0.0946	0.0306	-0.1201	0.0906
Kugelsto	0.5281	0.0946	1.0000	0.7442	0.0749	0.1376
Speerwur	0.2535	0.0306	0.7442	1.0000	-0.1255	0.0076
Weitspru	0.2727	-0.1201	0.0749	-0.1255	1.0000	0.5297
Kurzstre	0.0827	0.0906	0.1376	0.0076	0.5297	1.0000

1.) Zuerst werden die Kommunalitäten h² für jede Variable i geschätzt. Dafür werden entweder
  a) die multiplen Bestimmtheitsmaße R^2 verwendet - oder
  b) wie Rao auch vorschlägt. die Werte 0.5 verwendet.

In unserem Beispiel werden die multiplen Bestimmtheitsmaße R² eingesetzt

Eingesetzte Werte fuer Kommunalitaeten
(Multiple Bestimmtheitsmasse)

Marathon	0.5325
Rad	0.3010
Kugelsto	0.6951
Speerwur	0.6015
Weitspru	0.4609
Kurzstre	0.3659

2.) Die Einzelreste u_i^2 ergeben sich dann aus

$$u_i^2 = 1 - h_i^2$$

3.) Für die u_i^2 wird die Wurzel und dann der Kehrwert berechnet. Es entsteht u_i^{-1}

In unserem Beispiel ist dies

Vektor: U hoch -1
1.4625 1.1961 1.8110 1.5840 1.3620 1.2558

4.) Aus den u_i^{-1} und u_i^2 wird eine m * m Diagonalmatrix $\mathbf{U}^{-1}$ und $\mathbf{U}^2$ gebildet.

5.) Es wird die "Kanon.Fakt.Matrix" gebildet.

$U^{-1}\,(R - U^2)\,U^{-1}$

```
Iteration 0
Kanon.Fakt.Matrix:  U hoch -1 * (R-U hoch 2) * U hoch -1

         Marathon      Rad   Kugelst   Speerwu   Weitspr   Kurzstr
Marathon   1.1389   0.6953    1.3987    0.5873    0.5432    0.1519
Rad        0.6953   0.4307    0.2049    0.0580   -0.1956    0.1361
Kugelsto   1.3987   0.2049    2.2798    2.1350    0.1849    0.3129
Speerwur   0.5873   0.0580    2.1350    1.5092   -0.2707    0.0152
Weitspru   0.5432  -0.1956    0.1849   -0.2707    0.8550    0.9061
Kurzstre   0.1519   0.1361    0.3129    0.0152    0.9061    0.5771
```

6.) Für die so entstandene Matrix werden p Eigenwerte und normierte Eigenvektoren berechnet. Wir bezeichnen die Matrix der normierten Eigenvektoren mit **C** und die Diagonalmatrix der Eigenwerte mit **D**.

```
Iteration 0
Normierte Eigenvektoren aus Kanon.Fakt.Matrix je Faktor

         Faktor 1 Faktor 2 Faktor 3
Marathon   0.4014   0.3200  -0.5616
Rad        0.1057   0.0859  -0.6554
Kugelsto   0.7209  -0.0622   0.1343
Speerwur   0.5436  -0.3434   0.2832
Weitspru   0.0677   0.7039   0.2703
Kurzstre   0.0887   0.5224   0.2897

Eigenwerte aus Kanon.Fakt.Matrix größer 0
4.7544  1.8663  0.9799
```

7.) Nun muß entschieden werden, ob die Faktorenanalyse beendet werden soll oder ob das Iterationsverfahren eingeschlagen werden soll. Rao empfiehlt zu iterieren. Wird nicht iteriert, dann ist zu Punkt 12 zu springen.

8.) Für die Iteration muß eine feste Zahl von Faktoren bestimmt werden. Hier gibt es zwei Möglichkeiten:

a) Rao schlägt vor, den Lawley-Test der Faktorensignifikanz anzuwenden und dann nur für die signifikanten Faktoren zu iterieren. Dieses Verfahren hat den Nachteil, dass in späteren Iterationszyklen sich diese Zahl ändern kann, was zu nicht übersehbaren Änderungen führt. Wird in einem Rechnerprogramm dieses Verfahren gewählt, dann sollte man sich für jeden Iterationszyklus die Zahl der signifikanten Faktoren ausdrucken lassen.

b) Harris (1962, S. 262) schlägt vor, für die Iteration jene Faktoren beizubehalten, die bei der 1. Analyse eine kanonische Korrelation größer 0 besitzen. Das sind jene, deren Eigenwert bei der 1. Analyse größer .0. Nach Schluß der Analyse werden dann die statistisch signifikanten Faktoren bestimmt. Dieses Verfahren ist in unserem Computerprogramm realisiert. Es wird auch für die nachfolgenden Rechenschritte zugrunde gelegt.

9.) Die neuen u_i^{-1} werden dadurch berechnet, dass die Kommunalitäten h² der Lösung des 1. Zyklus berechnet und von 1.0 subtrahiert werden. Für die Variable i ergibt sich h² gemäß

$$h_i^2 = u_{ii}^2 \sum_{j=1}^{p} \lambda_j^* \cdot c_{ij}^2$$

wobei p = Zahl der extrahierten Faktoren. Der Index j kennzeichnet die Faktoren.

10.) Mit den neuen u_i^{-1} wird zu Punkt 5 zurückgesprungen und der nächste Zyklus gerechnet.

11.) Die Iteration wird so lange fortgesetzt, bis zwei aufeinanderfolgende u_i^{-1} sich sehr ähnlich sind (etwa .0005 Unterschied) oder eine vorgegebene Zahl von Zyklen erreicht ist.

```
Abbruch nach 10 Iterationen
maximale Differenz=0.0095  vorgebener Schwellenwert=0.001
Eigenwerte aus Kanon.Fakt.Matrix groesser 0
10.1467  4.5037  2.1517
```

12.) Nach dem Abbruch der Iteration wird die Matrix der Faktorladungen **F** berechnet

$$\mathbf{F} = \mathbf{U} \cdot \mathbf{C} \cdot \mathbf{D}^{\frac{1}{2}}$$

Matrix der Faktorladungen

	Faktor 1	Faktor 2	Faktor 3
Marathon	0.6639	0.3527	-0.4755
Rad	0.1887	0.0424	-0.5018
Kugelsto	0.9133	-0.1013	0.1024
Speerwur	0.7487	-0.3675	0.2307
Weitspru	0.1491	0.8251	0.2746
Kurzstre	0.1561	0.4877	0.3058

13.) Die reproduzierte Korrelationsmatrix **F*F'** und die Residualmatrix **R - F*F'** werden ermittelt
Das geschieht in Almo nur, wenn Zwischenergebnisse angefordert werden. Danach wird durch den Bartlett-Test überprüft, ob die Residualmatrix noch signifikant ist. Gewünscht ist, dass sie so minimale Werte besitzt, dass sie nicht mehr signifikant ist

Reproduzierte Matrix nach 3 Faktoren (Diagonale =Kommunalitaeten)

	Marathon	Rad	Kugels t	Speerwu	Weitspr	Kurzstr
Marathon	0.7913	0.3789	0.5219	0.2577	0.2594	0.1302
Rad	0.3789	0.2893	0.1167	0.0099	-0.0747	-0.1033
Kugelsto	0.5219	0.1167	0.8549	0.7447	0.0807	0.1245
Speerwur	0.2577	0.0099	0.7447	0.7488	-0.1283	0.0082
Weitspru	0.2594	-0.0747	0.0807	-0.1283	0.7784	0.5096
Kurzstre	0.1302	-0.1033	0.1245	0.0082	0.5096	0.3557

Residualmatrix nach 3 Faktoren (Diagonale =Einzelreste)

	Marathon	Rad	Kugels t	Speerwu	Weitspr	Kurzstr
Marathon	0.2087	0.0186	0.0062	-0.0043	0.0133	-0.0475
Rad	0.0186	0.7107	-0.0221	0.0207	-0.0454	0.1939
Kugelsto	0.0062	-0.0221	0.1451	-0.0004	-0.0057	0.0131
Speerwur	0.0043	0.0207	-0.0004	0.2512	0.0028	-0.0006
Weitspru	0.0133	-0.0454	-0.0057	0.0028	0.2216	0.0201
Kurzstre	0.0475	0.1939	0.0131	-0.0006	0.0201	0.6443

```
Signifikanz der Residualmatrix nach 3 Faktoren
Bartlett-Test der Sphaerizitaet
Determ.=0.9551  Chi-Quad.=2.1225  df=15  p=0.9998
```

14. Der Lawley-Test der Faktorensignifikanz wird durchgerechnet. Für jeden Faktor wird ein Chi-Quadrat-Wert berechnet.

```
Faktoren-Signifikanztest nach Lawley:
-------------------------------------
Chi-Quadrat-Werte je Faktor
   39.9024     15.0811      9.3653

Freiheitsgrade
      9           4           0

Wahrscheinlichkeit (1-p)*100, dass Faktor  signifikant
   99.9948     99.4927      0.0000

Beachte: In die Zahl der signifikanten Faktoren muss der
erste nicht mehr signifikante Faktor miteinbezogen werden
 - sofern ein solcher von Almo noch ausgegeben wird.
Ist schon der 1. nicht signifikant, dann ist dieser signifikant (!)
```

A1.2 Die Alpha-Faktorenanalyse

In der Testtheorie wird allgemein der Cronbach'sche Alpha-Koeffizient als der beste Zuverlässigkeitskoeffizient betrachtet. Er wird nach folgender Formel berechnet:

$$\alpha = \frac{m}{m-1} \left(1 - \frac{V}{\Sigma V}\right)$$

Bohmstedt (1969) zeigt, wie diese Formel auf die Kovarianzmatrix einer "Batterie" von Variablen (wir werden im Folgenden der Einfachheit halber von "Testbatterie" sprechen) angewendet werden kann. Dabei ist

V = Summe der Varianzen (= Summe der Diagonalglieder der Kovarianzmatrix)
ZV= Summe der Varianzen und Kovarianzen (= gesamte Matrixsumme)

Folgender Gedankengang liegt nun nahe: Hinter den Items steht (so wollen wir vorerst vereinfachend annehmen) ein gemeinsamer Faktor. Anstelle von V müssen wir dann die Varianz der Items, die durch den Faktor erklärt wird, einsetzen und anstelle von ZV, die Gesamtvarianz zwischen den Items.

Anstelle von V verwenden wir demzufolge die Diagonalmatrix der Kommunalitäten, wir nennen sie **H**, und anstelle von ZV die gesamte Kovarianz - bei standardisierten Variablen, die gesamte Korrelationsmatrix **R** minus der Diagonalmatrix der Einzelreste $\mathbf{U}^2$. Diese Matrizen müssen nun jedoch in skalare Größen überführt werden, die dann in die Alpha-Formel eingesetzt werden können.
Varianzen werden häufig durch quadratische Formen, also skalare Größen ausgedrückt. Wir verwenden als Vektor die Ladungen der Items auf dem gemeinsamen Faktor und bilden dann die beiden quadratischen Formen:

$$\mathbf{W}' \ \mathbf{H} \ \mathbf{W}$$

$$\mathbf{W}' \cdot (\mathbf{R} - \mathbf{U}^2) \cdot \mathbf{W}$$

Diese beiden Ausdrücke werden in die Alpha-Formel eingesetzt. Es entsteht:

$$(1) \quad \text{Alpha} = \frac{m}{m-1} \left(1 - \frac{\mathbf{W}' \cdot \mathbf{H} \cdot \mathbf{W}}{\mathbf{W}' (\mathbf{R} - \mathbf{U}^2) \mathbf{W}}\right)$$

Das ist die Ausgangsformel für die Alpha-Faktorenanalyse von Kaiser und Caffrey (1965).
Der gemeinsame den Variablen zugrunde liegende Faktor soll nun einen maximalen Alpha-Koeffizienten erzielen. Anders formuliert: Der Vektor **W**, mit dem die Korrelationsmatrix

(**R** - $\mathbf{U}^2$) und die Diagonalmatrix **H** prä- und postmultipliziert werden soll, muß so gewählt werden, dass Alpha ein maximaler Wert wird. Bei der „einfachen Faktorenanalyse" wird **W** so gewählt, dass die Varianz, die den l. Faktor erklärt, ein Maximum wird. Bei der „einfachen Faktorenanalyse" lautet die Gleichung, die es zu maximieren galt:

(2) Varianz = $\mathbf{W}' * \mathbf{R} * \mathbf{W}$

Eine Funktion wie in (1) oder (2) wird maximiert, indem ihre Ableitung gleich Null gesetzt wird. Bei (1) führt, ohne dass wir hier den Rechengang vorführen, (der interessierte Leser sei an die Originalarbeit von Kaiser und Caffrey verwiesen) zu der „charakteristischen Gleichung".

$$(3)\quad |\,(\mathbf{R} - \mathbf{U}^2) - \lambda\ \mathbf{H}^2\,|\ \mathbf{W} = \mathbf{O}$$

Nach verschiedenen Umformungen gelangt man schließlich zu der Matrix

$$(4)\quad \mathbf{M} = \mathbf{H}^{-1}\,(\mathbf{R} - \mathbf{I})\,\mathbf{H}^{-1} + \mathbf{I}$$

Wir bezeichnen **M** als "Alphamatrix". Sie ist die „Ausgangsmatrix", für die es gilt, die Eigenwerte λ und die jeweils dazugehörenden normierten Eigenvektoren zu finden. Dazu verwendet wird der "Tridiagonal-QR-Algorithmus" oder das v. Mises`schen Verfahren - was dann zu dem gesuchten Vektor **W** der Faktorenladungen führt.

Der Alpha-Koeffizient für den Faktor ist dann:

$$(5)\quad \text{Alpha} = \frac{m}{m-1}\left(1 - \frac{1}{\lambda_1}\right)$$

m = Zahl der Variablen λ_1 = 1. Eigenwert

Wir haben hier das Prinzip der Alpha-Faktorenanalyse der besseren Anschaulichkeit wegen nur für den 1. Faktor erklärt. Selbstverständlich gilt es, nachdem der 1. Faktor und seine Ladungen gefunden sind, orthogonal zu diesem, einen 2. Faktor mit dessen Ladungen zu finden - usw.
Der Vektor **W** muß also zur Matrix erweitert werden. Die Alpha-Faktorenanalyse liefert ein relativ klares Kriterium für die Zahl der zu extrahierenden Faktoren. Ein Faktor soll nur dann als bedeutsam betrachtet werden (wie Kaiser/Caffrey sagen: „may have a chance to find its place in the sun"), wenn er einen Alpha-Koeffizienten erzeugt, der größer 0 ist. Wenn wir Gleichung (5) betrachten, dann wird ersichtlich, dass dies nur geschieht, wenn der Eigenwert λ größer als 1 ist. Es sei davor gewarnt, das als die Lösung des Problems der Faktorensignifikanz zu betrachten. Es bleibt nach wie vor unberücksichtigt, dass die Korrelationsmatrix nicht Populations-, sondern Stichprobenwerte enthält.

Der Rechengang der Alpha-Faktorenanalyse

Wir werden im Folgenden anhand eines Beipiel die Rechenschritte durch Zahlen veranschaulichen
Wir verwenden die Daten der 6 Sportarten aus Abschnitt P30.2.3.4. Wird in Prog30m2.Msk die Optionsbox "weitere Optionen" geöffnet und "Zwischergebnisse" angefordert, dann gibt Almo diese Zahlen aus.
Die Ausgangs-Korrelationsmatrix ist

ursprüngliche Korrelationmatrix R

	Marathon	Rad	Kugelsto	Speerwu	Weitspr	Kurzstr
Marathon	1.0000	0.3975	0.5281	0.2535	0.2727	0.0827

Rad	0.3975	1.0000	0.0946	0.0306	-0.1201	0.0906
Kugelsto	0.5281	0.0946	1.0000	0.7442	0.0749	0.1376
Speerwur	0.2535	0.0306	0.7442	1.0000	-0.1255	0.0076
Weitspru	0.2727	-0.1201	0.0749	-0.1255	1.0000	0.5297
Kurzstre	0.0827	0.0906	0.1376	0.0076	0.5297	1.0000

1.) Zuerst werden für alle m Variable die Kommunalitäten h^2 geschätzt. Kaiser und Caffrey empfehlen entweder

a) die multiplen Bestimmtheitsmaße R^2 zu verwenden oder
b) einfach die Werte 1.0 als Kommunalitäten (vorläufig) zu betrachten (und bei der Kommunalitäteniteration dann auszutauschen)

Almo verwendet 1.0 als Kommunalitätenschätzung. Der Benutzer könnte aber auch auf R_QUADRAT umstellen. Wir raten davon ab, da je nach Datenkonstellation eine nicht positiv semidefinite Matrix entsteht !!

Um die R^2 berechnen zu können, muß die Matrix **R** invertiert werden. Die Diagonalelemente dieser lnversen seien cií. Dann gilt

$$(1)\quad h_i^2 = R_i^2 = 1.0 - \frac{1.0}{c_{ii}}$$

Werden die R^2 verwendet, dann kann eine nicht positiv semidefinite Matrix entstehen. D. h. es werden auch negative Eigenwerte berechnet. Wie Guttman nachgewiesen hat, ergibt die Zahl der positiven Eigenwerte in diesem Falle eine Maximalzahl der gemeinsamen Faktoren. D. h. mehr gemeinsame Faktoren (sehr wohl aber weniger) können nicht in der Korrelationsmatrix „stecken“.

2.) Es wird die Wurzel und dann der Kehrwert der h^2 gebildet. Damit entsteht h^{-1}
a) Wurden die R^2 als Kommunalitätenschätzung verwendet. dann gilt:

$$h^{-1} = \frac{1}{R_i}$$

b) Bei der in Almo voreingestellten Kommunalität von 1.0 ist selbstverständlich

h^{-1} = 1.0

3.) Die in 2 ermittelten Werte werden als Hauptdiagonale in eine m*m Diagonalmatrix $\mathbf{H}^{-1}$ eingeschrieben.

4.) Die Matrix (**R - I**) wird sehr einfach dadurch gebildet, dass die Diagonalglieder von **R** gleich 0 gesetzt werden. Wir schreiben: **R.o**

5.) Die Matrix **R.o** wird nun prä- und postmultipliziert mit der Diagonalmatrix $\mathbf{H}^{-1}$. Es entsteht die Matrix **R***.

$$\mathbf{R}^* = \mathbf{H}^{-1} \cdot \mathbf{R}_{.0} \cdot \mathbf{H}^{-1}$$

6.) Zu **R*** muß nun die Einheitsmatrix **I** wieder hinzuaddiert werden. Das geschieht sehr einfach dadurch, dass die Diagonale von **R*** (die ja, auch nach der Multiplikation mit $\mathbf{H}^{-1}$, immer noch 0 ist) gleich 1.0 gesetzt wird.
Damit ist die Ausgangsmatrix, die wir als "Alphamatrix" bezeichnet haben,

$$H^{-1}\ (R - I)\ H^{-1}\ + I$$

gebildet.
Bei Kommunalitäten-Iteration Nr. 0 ist die Matrix noch identisch mit der Ausgangs-Korrelationsmatrix.
Nach der 5. und letzten Kommunalitäten-Iteration gibt Almo als Zwischenergebnis aus

```
"Alphamatrix":  H hoch -1 * (R-I) * H hoch -1 + I
bei Kommunalitäten-Iteration 5
```

	Maratho	Rad	Kugelst	Speerwu	Weitspr	Kurzstr
Marathon	1.0000	0.6914	0.8045	0.4608	0.4536	0.2013
Rad	0.6914	1.0000	0.1154	0.0445	-0.1599	0.1766
Kugelsto	0.8045	0.1154	1.0000	0.9490	0.0874	0.2348
Speerwur	0.4608	0.0445	0.9490	1.0000	-0.1747	0.0155
Weitspru	0.4536	-0.1599	0.0874	-0.1747	1.0000	0.9869
Kurzstre	0.2013	0.1766	0.2348	0.0155	0.9869	1.0000

7.) Die Eigenwerte und Eigenvektoren der Alphamatrix werden berechnet.

```
Kommunalitäten-Iteration 5

Eigenwert je Faktor:     2.7918      1.9408      1.2065      0.3746
Varianz je Faktor:       1.8496      1.2279      0.8487
```

8.) Benötigt werden nur diejenigen Eigenwerte, die größer 1.0 sind. Wir symbolisieren sie durch
$\lambda_1, \lambda_2 \ldots\ldots \lambda_p$
Die zu ihnen gehörenden Eigenvektoren seien
$c_1, c_2 \ldots\ldots\ldots c_p$

Wir bilden eine p * p Diagonalmatrix der Eigenwerte (Symbol: **D**) und eine m * p Matrix der normierten Eigenvektoren (Symbol: **C**).

Nach der 5. und letzten Kommunalitäten-Iteration gibt Almo als Zwischenergebnis aus

```
Kommunalitäten-Iteration 5

Eigenwerte der Alphamatrix (sortiert nach absoluter Groesse):
              2.7918  1.9408  1.2065

          Normierte Eigenvektoren je Faktor
```

	Faktor 1	Faktor 2	Faktor 3
Marathon	0.5517	0.0539	-0.3008
Rad	0.2638	0.0924	-0.8056
Kugelsto	0.5351	0.2617	0.2929
Speerwur	0.4077	0.4055	0.3833
Weitspru	0.2749	-0.6482	0.1465
Kurzstre	0.3130	-0.5791	0.0802

9.) Es wird der Spaltenvektor der „uneigentlichen“ Faktorladungen für den Faktor j berechnet.

$$f_j^* = c_j \sqrt{\lambda_j}$$

Es entsteht eine Faktorladungsmatrix **F*** mit m Zeilen und p Spalten
(p = Faktoren mit Eigenwerten größer 1.0). Als Matrixprodukt dargestellt ist **F***

$$F^* = C \cdot D^{\frac{1}{2}}$$

Das sind noch nicht die endgültigen Faktorladungen. Die h^{-1} -Werte. mit denen ja in die Korrelationsmatrix hineinmultipliziert wurde, müssen wieder herausgenommen werden. Die „eigentliche" Faktorladungsmatrix ist dann

$$F = H \cdot F^* = H \cdot C \cdot D^{\frac{1}{2}}$$

(Siehe Punkt I3.)

10.) Nun ist eine Entscheidung zu treffen: Soll die Alpha-Faktorenanalyse ohne oder mit Kommunalitäteniteration weitergeführt werden? Kaiser und Caffrey empfehlen die Iteration. Das ist auch Standard in Almo.
Wir wollen den Rechengang für die Iteration darstellen. Wird nicht iteriert, dann muß zu Punkt l3 gesprungen werden.
In unserem Beispiel ist eine Komplikation aufgetreten, die nicht selten ist. Almo gibt nach Iteration 6 die neuen Kommunalitäten (berechnet in Punkt 11) aus.

```
Kommunalitaeten-Iteration 6
Kommunalitaeten: 0.4319  0.7135  1.0026  0.6083  0.8721  0.3219
```

Darauf folgt die Warnung

```
                              ******************** WARNUNG
                        Kommunalitaet groesser Diagonalglied (1.0)
  aufgetreten bei Iteration Nr. 6. Iterationszahl auf 5 herabgesetzt
                        Es muss nochmals von Neuem iteriert werden

                              ******************** WARNUNG
                                        Loesung problematisch !!!!
     Rechnen Sie besser eine Analyse ohne Kommunalitaeten-Iteration
```

Die Kommunalität der 3. Variablen hat bei der Iteration von 5 zu 6 den Wert 1.0 überschritten.

ll.) Es werden nun neue Kommunalitätenschätzungen aus der gewonnenen Faktorladungsmatrix F* berechnet. Die neue Kommunalität h_i^{*2} für die Variable i ist

$$h^{*2}_i = h_i^2 \cdot \sum_{j=1}^{p} c_{ij}^2 \cdot \lambda_j = h_i^2 \cdot \sum_{j=1}^{p} f_j^{*2}$$

h_i^{*2} = neue Kommunalität

h_i^2 = vorherige Kommunalität

f_j^{*2} = siehe Punkt 9

p = Zahl der extrahierten Faktoren

Rechnerisch geht man am besten so vor. dass man die Elemente der Faktorladungsmatrix **F*** einzeln für sich quadriert. Die dabei entstandenen Werte werden in jeder Reihe nach hinten hin aufaddiert. Dieser Summenwert wird dann mit der alten Kommunalität h² multipliziert. Damit gewinnen wir für jede Variable die neue Kommunalität h*².

12.) Nun wird zu Punkt 2 zurückgesprungen, d. h. es wird die Wurzel und der Kehrwert der neuen Kommunalitäten berechnet. Danach werden die folgenden Schritte bis Punkt 10 durchgeführt. Hier ist nun jedes mal die Entscheidung zu treffen, ob die Iteration fortgesetzt oder abgebrochen werden soll. Als Entscheidungskriterien kommen hier in Frage:

a) wenn sich die neuen Kommunalitäten von denen des vorhergehenden Interationszyklus nur noch wenig unterscheiden (voreingestellt: nur noch um .001), dann kann abgebrochen werden. Almo liefert dann ein korrektes Ergebnis. Der Almo-Benutzer kann diesen Schwellenwert in der Optionsbox "Kommunalitätenschätzung" veränern.
b) Wenn eine maximale Zahl von Iterationen durchlaufen ist, dann wird, sofern Kriterium a nicht schon vorher erfüllt wurde, abgebrochen. Der Almo-Benutzer kann die Zahl der Iterationen in der Optionsbox "Kommunalitätenschätzung" einstellen.
c) Almo bricht selbsttätig schon nach wenigen Iterationen ab, wie in unserem Beispiel, weil Kommunalitäten größer 1.0 entstanden sind.

13.) Nach dem Abbruch der Iteration werden die „eigentlichen" Faktorladungen berechnet. Die in die Korrelationsmatrix hineinmultiplizierten h^{-1} -Werte, werden sozusagen wieder herausgenommen. Die „eigentliche" Faktorladungsmatrix ist dann

$$\mathbf{F = H \cdot F^* = H \cdot C \cdot D^{\frac{1}{2}}}$$

Almo gibt standardmäßig das Hauptergebnis der Alpha-Faktorenanalyse aus:

Matrix der Faktorladungen

	Faktor 1	Faktor 2	Faktor 3
Marathon	0.6254	0.0509	-0.2242
Rad	0.3734	0.1091	-0.7496
Kugelsto	0.8649	0.3527	0.3112
Speerwur	0.5523	0.4580	0.3413
Weitspru	0.4070	-0.8002	0.1426
Kurzstre	0.3168	-0.4887	0.0534

14.) Es können nun noch eine Reihe von Koeffizienten berechnet werden:

a) Die Varianzen je Faktor.
Die Elemente der „eigentlichen" Faktormatrix **F** werden einzeln für sich quadriert. Dann wird spaltenweise summiert. Diese Spaltensumme ist die Varianz je Faktor.
Almo gibt standardmäßig als Ergebnis der Alpha-Faktorenanalyse aus:

```
Varianz je Faktor
    1.8496      1.2279      0.8487

Prozent der Varianz
   30.8274     20.4643     14.1451

Zu erklaerende Gesamtvarianz=        6.0000
Durch 3 Faktoren erklaerte Varianz=  3.9262
Prozentsatz der erklaerten Varianz= 65.4368
```

b) Die Kommunalitäten je Variable entstehen, indem die - wie unter ll beschriebenen - quadrierte Faktorladungsmatrix zeilenweise aufaddiert wird. Die Zeilensummen sind die Kommunalitäten je Variable.
Almo gibt standardmäßig das Hauptergebnis der Alpha-Faktorenanalyse aus:

Kommunalitaeten je Variable

Marathon	0.4440
Rad	0.7132
Kugelsto	0.9692
Speerwur	0.6313
Weitspru	0.8263
Kurzstre	0.3421

c) Die Alpha-Koeffizienten je Faktor. Ziel der Alpha-Faktorenanalyse ist es ja, die Faktorladungen so zu wählen, dass ein maximaler Alpha-Koeffizient je Faktor entsteht. Es ist also sinnvoll, die Alpha-Koeffizienten der Faktoren zu berechnen. Das geschieht nach der Formel:

$$\text{Alpha}_j = \frac{m}{m-1} \left(1 - \frac{1}{\lambda_j}\right)$$

wobei m die Zahl der Variablen und λ_j der Eigenwert des Faktors j aus der Alphamatrix ist (wie er in Punkt 8 berechnet wurde).
Wird die Alpha-Faktorenanalyse dazu verwendet, um die Eindimensionalität einer Testbatterie zu überprüfen, so gibt der Alpha-Koeffizient des 1. Faktors den Zuverlässigkeitskoeffizienten Alpha für die auf Eindimensionalität bereinigte Testbatterie an. Das ist ein sehr wertvolles Maß.
Almo gibt standardmäßig als Ergebnis der Alpha-Faktorenanalyse aus:

```
Alpha-Koeffizient je Faktor
    0.6418      0.4848      0.1711
```

d) Wurde iteriert, dann ist es sinnvoll, sich die Differenz zwischen dem letzten und vorletzten Kommunalitätenvektor ausgeben zu lassen. Man kann dann den Grad der Konvergenz beurteilen.
Almo gibt standardmäßig als Ergebnis der Alpha-Faktorenanalyse aus:

Kommunalitaeten-Iterationen: 5
Letzte Iterationsdifferenz (je Variable)

Marathon	V1	0.0164
Rad	V2	0.0045
Kugelsto	V3	0.0334
Speerwur	V4	0.0259
Weitspru	V5	0.0411
Kurzstre	V6	0.0249

Interessant ist es, zu überprüfen, ob das Iterieren konvergiert, d.h. ob die Kommunalitäten-Differenz je Variable gegen 0 strebt.

Konvergenz der Kommunalitaeten-Iterationen

Kommunalitaeten-Differenz	Maratho	Rad	Kugelst	Speerwu	Weitspr	Kurzstr
von Iteration -1 zu 0	0.273	0.131	0.100	0.143	0.178	0.317
von Iteration 0 zu 1	0.133	0.075	0.028	0.081	0.071	0.153
von Iteration 1 zu 2	0.070	0.042	0.008	0.052	0.014	0.081
von Iteration 2 zu 3	0.039	0.023	0.024	0.037	0.016	0.049
von Iteration 3 zu 4	0.024	0.012	0.031	0.030	0.032	0.033
von Iteration 4 zu 5	0.016	0.004	0.033	0.026	0.041	0.025
konvergiert?	ja	ja	nein	ja	nein	ja
mit Toleranz von 0.009	ja	ja	nein	ja	nein	ja

Bei den Variablen 1, 2, 4, 6 werden von oben nach unten die Differenzwerte kleiner. Sie streben gegen 0. Es besteht Konvergenz. Das Iterieren war erfolgreich. Nicht jedoch bei `Kugelstoss` und `Weitsprung`. Bei beiden steigt die Differenz schon bei der Iteration von Schritt 2 zu 3 auf. Auch kann mit einem Toleranzwert von +0.009 zum jeweils vorausgehenden Wert keine Konvergenz konstatiert werde.

Zum Rechengang der Alpha-Faktorenanalyse muß folgendes angemerkt werden:

1) Bei der Kommunalitäteniteration ist es wichtig, dass immer mit der gleichen Zahl von Faktoren gearbeitet wird, da sonst eine Konvergenz unmöglich wäre.

2) Die Iteration kann in seltenen Fällen zu Kommunalitäten h^2 führen, die größer als 1.0 sind. Dieser Fall tritt eher auf, wenn als Kommunalitätenschätzungen 1.0 vorgegeben wurde (was standardmäßig in Almo geschieht) und seltener wenn R_QUADRAT verwendet werden. Kaiser und Caffrey meinen, dass das nicht so schlimm wäre. Hier wird empfohlen, die letzte Lösung vor derjenigen, die zu Kommunalitäten größer.l.0 geführt hat, als die bestmögliche zu erachten. In unserem Computer-Programm wird so verfahren.

A1.3 Die Image-Faktorenanalyse

Wenn m Variable miteinander korrelieren, dann läßt sich eine einzelne, herausgegriffene Variable l als durch die anderen m-l Variablen bestimmt betrachten. Diese Bestimmung braucht keine solche im Sinne der Kausalität zu sein. Wir können eine multiple Regressionsgleichung aufstellen:

$$X_1 = a \cdot X_2 + b \cdot X_3 + c \cdot X_4 + \ldots\ldots + e$$

wobei die Werte der Variablen X1 durch die Werte der anderen Variablen X2, X3 ... Xm bestimmt werden (nachdem die Gewichtszahlen a, b, c . . . bekannt sind). Dieses Bestimmen gelingt nicht vollständig. Es bleibt ein Fehleranteil e. Wir können nun auch schreiben:

$$(1)\quad X_1 = p + e$$

p ist der durch die anderen m-l Variablen prognostizierte Wert der Variablen 1, e ist der nicht prognostizierte Anteil.
Guttman nennt nun "p" das „Image“ der Variablen l und "e" das „Anti-Image“ der Variablen l.
Man kann für zwei Variable deren durch die jeweils anderen Variablen prognostizierten Wert - deren Image - durch multiple Regression bestimmen. Jeder einzelne Probandenwert wird dann für Variable l aufgespalten in einen Imagewert pl und einen Anti-Imagewert e1 - fiir die Variable 2 entsprechend in p2 und e2.
Wir werden im Folgenden das Anti-Image nicht durch "e" sondern durch "a" symbolisieren

a = e = Anti-Image

Die p-Werte der beiden Variablen lassen sich korrelieren. Es entsteht die Kovarianz g12 (bzw. bei standardisierten Variablen die Korrelation g12).
Auch die e- bzw. a-Werte lassen sich korrelieren. Es entsteht die Kovarianz (bzw. bei standardisierten Werten die Korrelation) a12. Nun gilt die (hier nicht beweisbare) „fundamentale Gleichheit“:

(2) r12 = g12 - a12

Die Korrelation zwischen den beiden Variablen l und 2 ist gleich der Differenz der Korrelation des Image und des Anti-Image der beiden Variablen. Das Verfahren läßt sich nun für alle Variablen durchführen, d. h. wir können für alle Variablen deren Werte in einen p- und einen e-Anteil zerlegen und dann jeweils die p- und e-Anteile interkorrelieren. Auf diese Weise erhalten wir zwei Matrizen, die Matrix der Image-Korrelationen - wir nennen sie kurz Image-Matrix und symbolisieren sie durch **G** - und die Matrix der Anti-Image Korrelationen - wir nennen sie kurz Anti-Image-Matrix und symbolisieren sie durch **A**. Gemäß der oben erwähnten „fundamentalen Gleichheit“ ist dann

(3) $\mathbf{R = G - A}$
(3a) $\mathbf{G = R + A}$

In unserem Computer-Programm wird **G** als "Imagematrix" oder "unskalierte Imagematrix" bezeichnet
Die Korrelationsmatrix **R** entsteht dadurch, dass die Anti-Image-Matrix **A** von der Image-Matrix subtrahiert wird. Es ist nun nicht sinnvoll, die Matrix der Anti-Images in eine Faktorenanalyse miteinzubeziehen (da sie ja - etwas unscharf formuliert - die Anteile der Variablen enthält, die fremdbestimmt sind). Die Faktorenanalyse ist auf die Image-Matrix **G** als Ausgangsmatrix zu richten. Die Image-Matrix **G** braucht nicht auf dem umständlichen Weg der m-fachen Durchführung einer multiplen Regression berechnet zu werden. Sie kann, mathematisch elegant, (ohne dass wir das hier beweisen) in folgender Weise gewonnen werden:

(4) $\mathbf{G = R + Q^{-1}\ R^{-1}\ Q^{-1} - 2\ Q^{-1}}$

wobei die Anti-Image-Matrix **A**

(4a) $\mathbf{A} = \mathbf{Q^{-1}\ R^{-1}\ Q^{-1} - 2\ Q^{-1}}$.

Ein einzelnes Glied der Imagematrix gij ist dann

gij = rij + aij

R = Korrelationsmatrix mit 1.0 in der Diagonalen
$\mathbf{R^{-1}}$ = Inverse dieser Matrix
$\mathbf{Q^{-1}}$ = das ist eine Diagonalmatrix, in deren Diagonale die Werte
$1\text{-}R^2$ stehen, d. h. die Werte $1/c_{ii}$ wobei c_{ii} das entsprechende
Diagonalglied aus der invertierten Korrelationsmatrix $\mathbf{R^{-1}}$ ist.
gij = das Glied ij aus der Image-Matric **G**
rij = das Glied ij aus der ursprünglichen Korrelationsmatrix **R**
aij = das Glied ij aus der Anti-Image-Matric **A**

Die **G**-Matrix, die wir ja der Faktorenanalyse unterziehen wollen, besitzt eine wertvolle Eigenschaft: Sie ist positiv semidefinit. Alle m Eigenwerte und Eigenvektoren sind berechenbar. Wir wollen untersuchen, welche Werte in der Hauptdiagonalen der Matrix **G** stehen. Die drei Diagonalglieder der drei Teil-Matrizen sind

Teilmatrizen	Diagonalglied
R	1.0
$\mathbf{Q^{-1} \cdot R^{-1} \cdot Q^{-1}}$	$\frac{1}{c_{ii}} \cdot c_{ii} \cdot \frac{1}{c_{ii}}$
$2\ \mathbf{Q^{-1}}$	$2\ \frac{1}{c_{ii}}$

Das Diagonalglied der Imagematrix ist dann

$$(5)\quad 1.0 + \frac{1}{c_{ii}} \cdot c_{ii} \cdot \frac{1}{c_{ii}} - 2\,\frac{1}{c_{ii}}$$

Dieser Ausdruck ergibt $1 - c_{ii} = R^2_{ii}$
d. h. in der Hauptdiagonalen der **G**-Matrix stehen die multiplen Bestimmtheitsmaße R^2ii.

Wir können sagen, die **G**-Matrix entsteht aus der Korrelationsmatrix **R** und der Anti-Image-Matrix **A** oder äquivalent: die Korrelationsmatrix **R** entsteht aus der Imagematrix **G** minus der Anti-Imagematrix **A.** Wir wiederholen die obigen Gleichungen 3 und 3a

(6) **R = G - A**

(6a) **G = R + A**

Es gilt nun der wichtige Satz, (den wir hier auch nicht beweisen): Ist der Rang der Korrelationsmatrix, d. h. ist die Zahl gemeinsamer Faktoren, die in **R** „stecken", kleiner als die Zahl der Variablen, dann sind die Elemente der Anti-Image-Matrix **A**, die außerhalb der Diagonalen stehen, nahe 0. Nun sind zwar Korrelationsmatrizen, da sie Stichproben- und nicht Populationswerte enthalten, immer vom Rang m, so dass die **A**-Matrix außerhalb der Diagonalen kleine, mehr oder weniger von Null verschiedene Werte besitzt.

Das führt uns nun zu der Formulierung, dass die Matrix **G** eine normale Korrelationsmatrix **R** ist, die in der Diagonalen als Kommunalitätenschätzung R^2ii als Werte enthält und deren außerhalb der Diagonale liegende Werte korrigiert wurden - mit dem Ergebnis, dass die Matrix positiv semidefinit wurde. Wir können also, wenn wir die hinter dem Guttman-Verfahren stehende Theorie nicht berücksichtigen, dieses Verfahren in die „einfache Faktorenanalyse" einordnen und als eine Methode bezeichnen, die der Korrelationsmatrix die Eigenschaften erhält, positiv semidedefinit zu sein - obwohl R^2ii -Werte als Kommunalitätenschätzungen verwendet werden. Das hat zur Folge, dass aus der Imagematrix so viele Faktoren extrahiert werden können, wie Variable vorhanden sind. Dem Forscher stellt sich dadurch die Frage: Wie viele Faktoren sind tatsächlich bezüglich des Objekts, das ich untersuche, relevant?

Die relevante Faktorenzahl
Almo verwendet folgende Automatik. Es ermittelt, die Zahl der Eigenwerte größer 1.0 aus der Korrelationsmatrix **R** (mit 1.0 in der Diagonalen - nicht mit den R_Quadrat in der Diagonalen) und fixiert diese als die Zahl der Faktoren, die aus der Imagematrix **G** zu extrahieren sind. Tatsächlich faktorisiert Almo nicht G sondern, wie wir noch zeigen werden die skalierte Imagematrix G* , was jedoch keinen Unterschied macht. Wenn der Forscher sich dieser Automatik nicht unterwerfen will, dann muß er in Prog30m2 die Optionsbox "Faktoren" bzw. im Bootstrap-Programm Prog30ml "Faktoren und Kommunalitäten" öffnen und eine Faktorenzahl einsetzen.

Ein Beipiel:
Wir verwenden die Daten der 6 Sportarten aus Abschnitt P30.2.3.4. Wird in Prog30m2.Msk die Optionsbox "weitere Optionen" geöffnet und "Zwischergebnisse" angefordert, dann gibt Almo die (unskalierte) Imagematrix **G** aus.

ursprüngliche Korrelationmatrix R

	Marathon	Rad	Kugelsto	Speerwu	Weitspr	Kurzstr

Marathon	1.0000	0.3975	0.5281	0.2535	0.2727	0.0827
Rad	0.3975	1.0000	0.0946	0.0306	−0.1201	0.0906
Kugelsto	0.5281	0.0946	1.0000	0.7442	0.0749	0.1376
Speerwur	0.2535	0.0306	0.7442	1.0000	−0.1255	0.0076
Weitspru	0.2727	−0.1201	0.0749	−0.1255	1.0000	0.5297
Kurzstre	0.0827	0.0906	0.1376	0.0076	0.5297	1.0000

(unskalierte) Imagematrix G
(Diagonale = multiple Bestimmtheitsmasse)

	Marathon	Rad	Kugelsto	Speerwu	Weitspr	Kurzstr	Multiple Bestimmtheits-maße
Marathon	0.5325	0.1098	0.3385	0.3221	0.0670	0.2355	0.5325
Rad	0.1098	0.3010	0.1658	0.0337	0.1118	−0.0995	0.3010
Kugelsto	0.3385	0.1658	0.6951	0.4895	0.0938	0.0571	0.6951
Speerwur	0.3221	0.0337	0.4895	0.6015	−0.0518	0.0218	0.6015
Weitspru	0.0670	0.1118	0.0938	−0.0518	0.4609	0.1916	0.4609
Kurzstre	0.2355	−0.0995	0.0571	0.0218	0.1916	0.3659	0.3659

Anti−Image−Matrix A
(Diagonale = 1.0−multiple Bestimmtheitsmasse)

	Marathon	Rad	Kugelst.	Speerwu	Weitspr.	Kurzstr.
Marathon	0.4675	−0.2877	−0.1896	0.0686	−0.2057	0.1528
Rad	−0.2877	0.6990	0.0712	0.0031	0.2319	−0.1901
Kugelsto	−0.1896	0.0712	0.3049	−0.2547	0.0188	−0.0805
Speerwur	0.0686	0.0031	−0.2547	0.3985	0.0737	0.0142
Weitspru	−0.2057	0.2319	0.0188	0.0737	0.5391	−0.3381
Kurzstre	0.1528	−0.1901	−0.0805	0.0142	−0.3381	0.6341

Werden die Werte der Korrelationsmatrix **R** mit den Werten der Anti-Imagematrix **A** addiert, dann entstehen daraus die Werte der Imagmatrix **G**. Wir zeigen dies für das Matrixelement (1,2) Marathon versus Rad

```
  r(1,2)     0.3975
+ a(1,2)    -0.2b77
-----------------
  g(1,2)     0.1098
```

Der Unterschied zwischen der ursprünglichen Korrelationsmatrix **R** und der Imagematrix **G** ist doch beträchtlich.

Matrix der Faktorladungen aus "normaler" Faktorenanalyse
(mit multiplen Bestimmtheitsmaßen als Kommunalitaeten)

		Faktor 1	Faktor 2	Faktor 3
Marathon	V1	0.6453	0.1710	−0.3742
Rad	V2	0.2441	0.0157	−0.5256
Kugelsto	V3	0.8596	−0.1405	0.1559
Speerwur	V4	0.6820	−0.3590	0.2611
Weitspru	V5	0.1835	0.7050	0.1530
Kurzstre	V6	0.2126	0.5743	0.1394

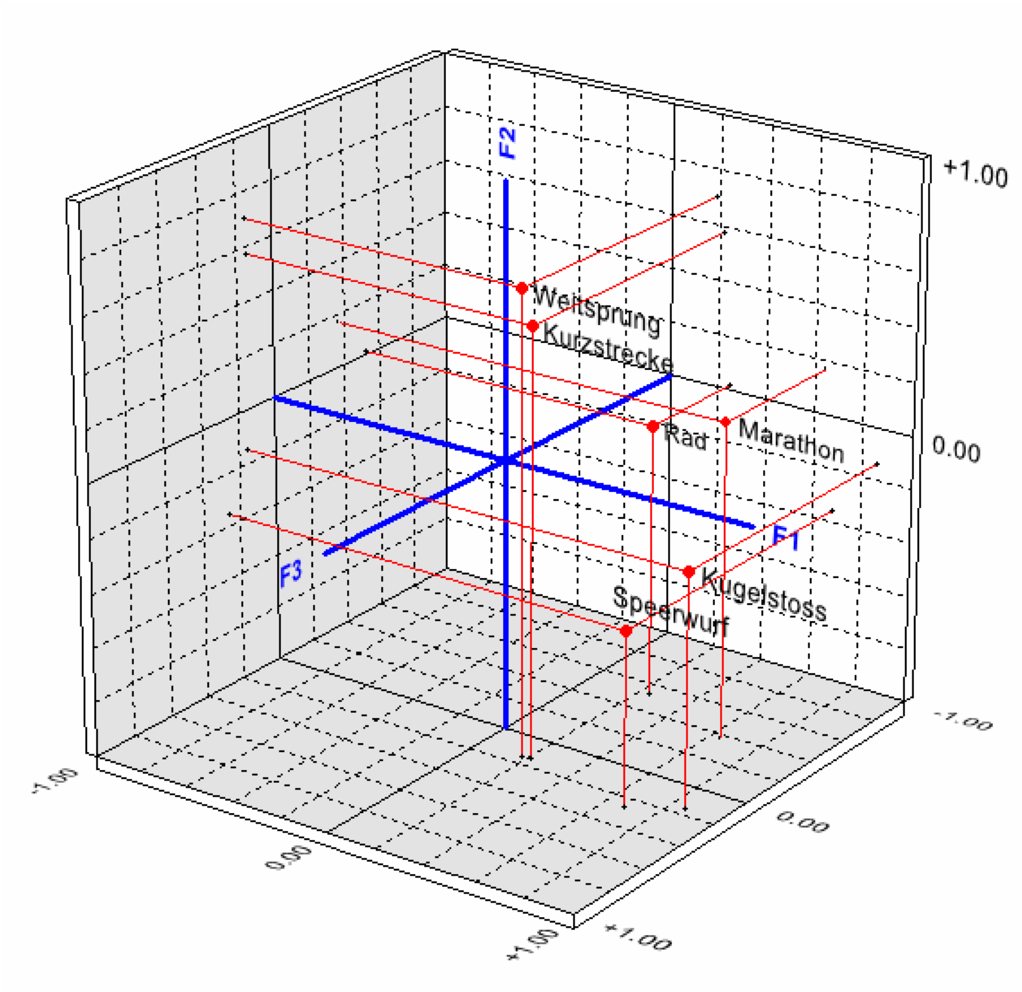
Matrix der Faktorladungen aus Image−Faktorenanalyse

		Faktor 1	Faktor 2	Faktor 3
Marathon	V1	0.5440	0.2412	−0.3233
Rad	V2	0.1752	0.0791	0.3194
Kugelsto	V3	0.7890	−0.0379	0.2073

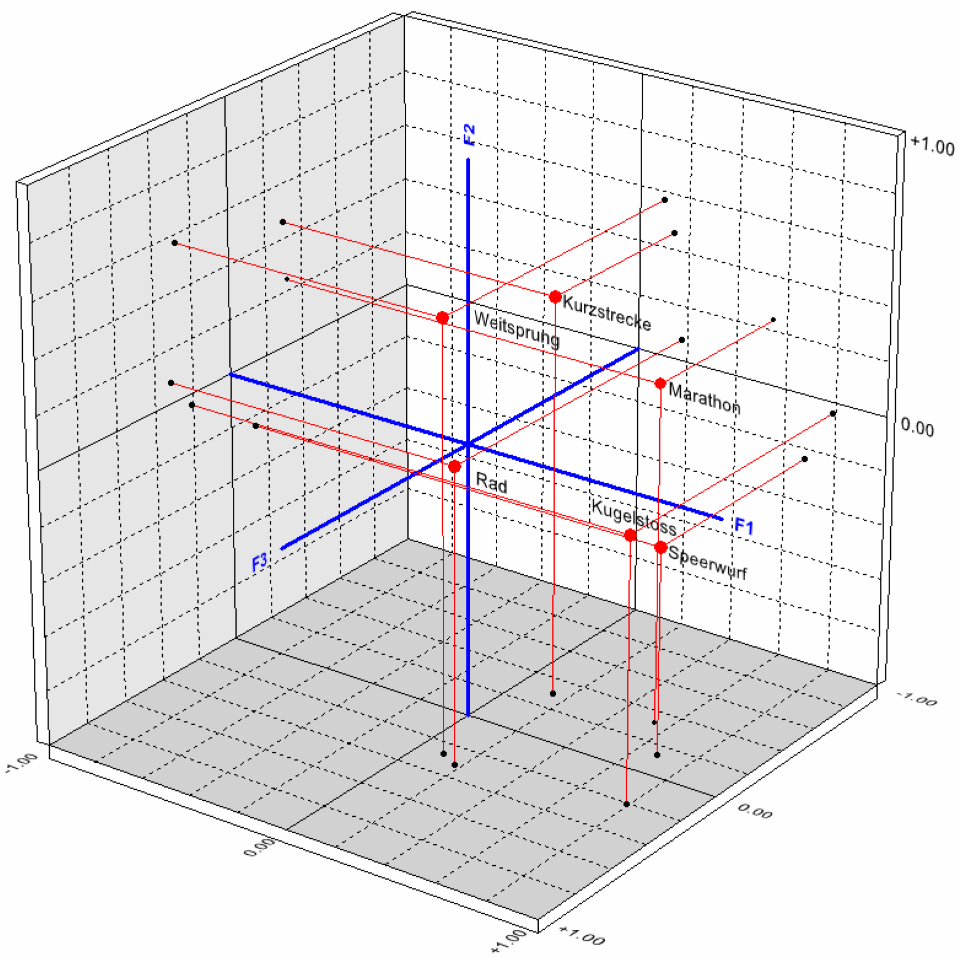
Speerwur	V4	0.6801	-0.2390	-0.1318
Weitspru	V5	0.0985	0.5697	0.2777
Kurzstre	V6	0.1400	0.4586	-0.2989

Die Faktorladungen differieren beträchtlich. Das wird bei der graphischen Darstellung deutlich

Grafik der Faktorladungen aus "normaler" Faktorenanalyse



Grafik der Faktorladungen aus Image-Faktorenanalyse



Von "Rad" und "Marathon" würde man erwarten, dass sie eine Gruppe bilden. Sie sind jedoch in der Achse F3 zu weit voneinander entfernt, wie man feststellen kann, wenn man die linke Seitenwand betrachtet.

Skalierte Imagematrix
C. W. Harris (1962) hat vorgeschlagen, die Imagematrix **G** gemäß folgender Gleichung zu skalieren

(7) $$\mathbf{G^* = Q^{\frac{1}{2}} \cdot G \cdot Q^{\frac{1}{2}}}$$

In unserem Computer-Programm wird **G*** berechnet und als "skalierte Imagematrix" bezeichnet. Die Matrix **G*** wird dann faktorisiert. Der dabei entstehende Vorteil ist, dass anstelle der Korrelationsmatrix die Kovarianzmatrix verwendet werden kann - ohne dass sich dadurch die Ergebnisse grundlegend ändern. Die "skalierte Imagematrix" **G*** ist dieselbe. Das gilt in Almo sogar für die Quadratsummenmatrix, allerdings muss für diese die Faktorenzahl vom Benutzer festgelegt werden. Eine Automatik mit der die relevante Faktorenzahl gefunden wird, wie sie oben beschrieben wurde und wie sie für die Korrelationsmatrix und die Kovarianzmatrix vorhanden ist, ist für die Quadratsummenmatrix nicht existent.

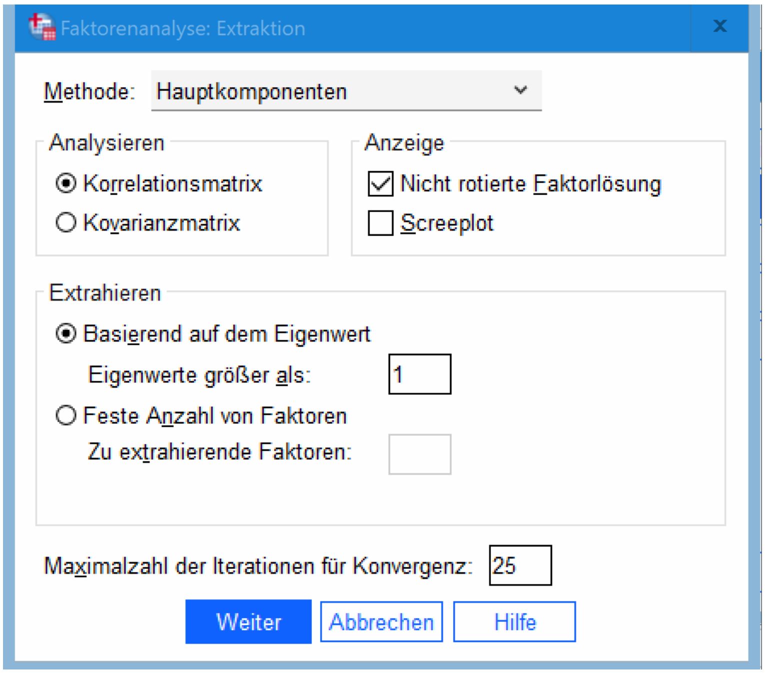
Anhang 2: Vergleich Almo mit SPSS

Wir verwenden die SPSS-Version 27 und rechnen im folgenden ein Beispiel mit der SPSS-Datei "HolzingerSwineford1939.sav" und für Almo mit "HolzingerSwineford 1939.fre". Beide Dateien befinden sich im Almo-Ordner TESTDAT. Die Datei wird kurz beschrieben im Almo-Dokumen 15a "Bootstrap bei Faktorenanalyse", Abschnitt P30.6.2.2.

Verwendet werden diese 6 Variablen

```
   Variablen-   Variablennamen
 nummer in Almo
 -------------  -----------------
          V14    t6_paragraph_comprehension
          V15    t7_sentence
          V17    t9_word_meaning
          ---------------------------------
          V22    t14_word_recognition
          V23    t15_number_recognition
          V25    t17_object_number
```

Wird in SPSS die Option "Extraktion" geöffnet dann sieht man folgendes



Betrachtet man die zentrale Eingabe-Box der SPSS-Faktorenanalyse, dann wird eine überraschende Einschränkung sichtbar: Der Benutzer kann die Kommunalitäten-Schätzung nicht beeinflussen. SPSS hat vermutlich im Bemühen, dem Benutzer die Eingabe möglichst leicht zu machen, auf diese Option verzichtet. Indirekt entscheidet sich der Benutzer für "1.0 als Kommunalitätenschätzung" wenn er als Methode "Hauptkomponenten" wählt und für "multiple Bestimmtheitsmaße (=R-Quadrat)" wenn er "Hauptachsen" anfordert.

A2.1 SPSS: Hauptkomponentenanalyse

Der Benutzer kann zwischen den beiden Vorgehensweisen "Basierend auf dem Eigenwert" und "Feste Anzahl von Faktoren" wählen

"Basierend auf dem Eigenwert"
Mit der abgebildeten Einstellung rechnet SPSS eine Hauptkomponentenanalyse für eine Faktorenanalyse mit diesen Einstellungen:

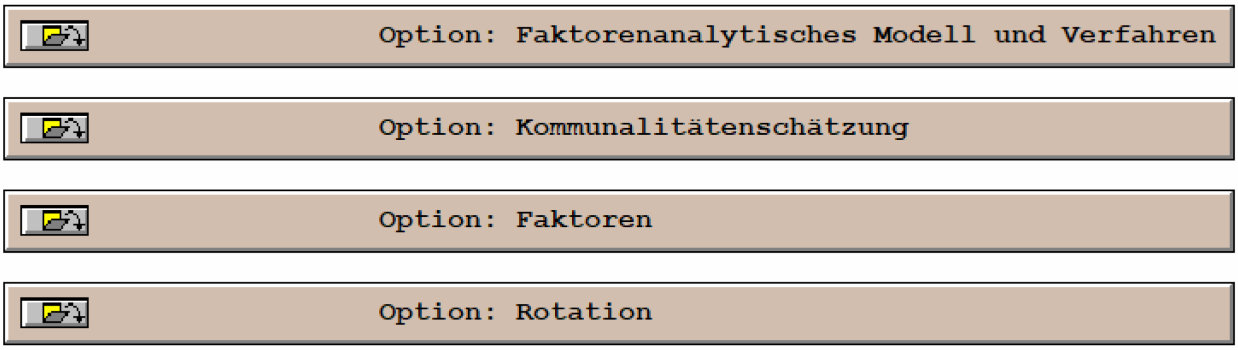
1. Als Kommunalitäten werden 1.0 in die Diagonale eingesetzt

2. Die Eigenwerte werden entsprechend aus der Korrelationsmatrix mit 1.0 in der Diagonalen berechnet. Punkt 1 und 2 sind charakteristisch für die Hauptkomponentenanalyse und können vom Benutzer nicht beeinflusst werden.
3. Für die eigentliche Faktorenanalyse werden dann die Faktoren extrahiert, deren Eigenwerte gemäß Punkt 2 größer 1.0 sind. Dieser Wert kann vom Benutzer verändert werden. Wird 0 eingesetzt, dann werden alle Faktoren extrahiert. D.h. es werden so viele Faktoren extrahiert wie Variable vorhanden sind. Die Faktorenanalyse kann mit einer Kommunalitäten-Iteration gerechnet werden. Am Unterrand der Box werden standardmäßig 25 Iterationszyklen angeboten. Wird 1 eingegeben, dann wird nicht iteriert.

"Feste Anzahl von Faktoren"
Maximal können soviele Faktoren angefordert werden wie Variable vorhanden sind. Als Kommunalitätenschätzung wird automatisch vom Programm 1.0 eingesetzt. Die Kommunalitäten können iteriert werden.

A2.2 Almo: Hauptkomponentenanalyse

In Almo wird durch nachfolgend abgebildete Optionsboxen gesteuert, welches Faktorenmodell gerechnet werden soll:



Die SPSS-Einstellungen für die Hauptkomponentenanalyse werden in Almo erreicht, wenn diese 4 Optionsboxen geschlossen bleiben. Dann gilt:

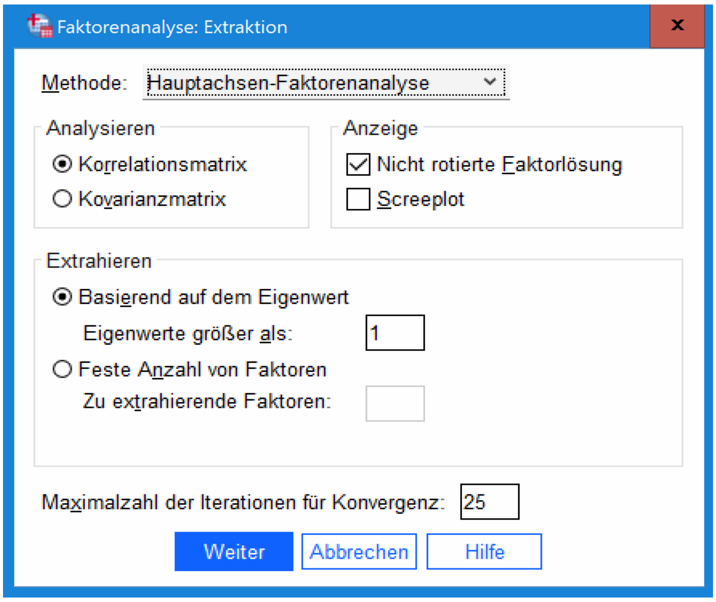
1. die Eigenwerte werden aus der Korrelationsmatrix mit 1.0 in der Diagonalen berechnet.
2. Für die Faktorenanalyse werden dann die Faktoren extrahiert, deren Eigenwerte größer 1.0 sind.
3. Eine Iteration der Kommunalitäten findet nicht statt.

zu 2. Soll eine andere Faktorenzahl extrahiert werden, dann muss die Optionsbox "Faktoren" geöffnet werden und die gewünschte Zahl eingetragen werden.

zu 3. Almo ermöglicht auch eine Analyse mit Kommunalitätenschätzung=1.0 und Kommunalitäten-Iterationen größer 0. Dazu muss die Optionsbox " Kommunalitätenschätzung" geöffnet werden.

Almo rotiert standarmäßig die Faktorladungsmatrix. Das kann unterdrückt werden durch Öffnen der Optionsbox "Rotation" und eine entsprechende Eingabe.

A2.3 SPSS: Hauptachsenanalyse



Bei der Hauptachsenanalyse setzt SPSS die multiplen Bestimmtheitsmaße (R_QUADRAT) als Kommunalitäten-Schätzung ein (siehe nachfolgende Ergebnis-Tabelle 1), rechnet jedoch zuerst sämtliche Eigenwerte für die Korrelationsmatrix mit Diagonalwerten=1 und 0 Iterationen. Es wird gewissermaßen eine "vorzeitig abgebrochene" Haupt-*komponenten*- analyse gerechnet (Tabelle 2, vordere Hälfte).

SPSS bietet dem Benutzer zwei Vorgehensweisen an:

"Basierend auf dem Eigenwert"
Voreingestellt ist "alle Faktoren mit Eigenwert größer 1.0. Gemeint sind jedoch die Eigenwerte aus der "vorzeitig abgebrochenen" Haupt-*Komponenten*-Analyse mit 1.0 als Kommunalitäten, nicht diejenigen, die sich aus der Haupt-*Achsen*-Analyse mit R-Quadrat als Kommunalität ergeben würden. Das ist eine nicht ganz folgerichtige Vorgehensweise, die sich aber weitgehend bewährt hat. In der Regel gelingt es damit eine sinnvolle, d.h. inhaltlich interpretierbare Zahl von Faktoren zu extrahieren.

SPSS liefert folgendes Ergebnis

Tabelle 1

Kommunalitäten

	Anfänglich	Extraktion
t6_paragraph_comprehension	,611	,721
t7_sentence	,625	,764
t9_word_meaning	,589	,686
t14_word_recognition	,245	,457
t15_number_recognition	,199	,368
t17_object_number	,167	,272

Extraktionsmethode: Hauptachsen-Faktorenanalyse.

Tabelle 2

Faktor	Anfängliche Eigenwerte Gesamt	% der Varianz	Kumulierte %	Summen von quadrierten Faktorladungen für Extraktion Gesamt	% der Varianz	Kumulierte %
1	2,591	43,182	43,182	2,278	37,972	37,972
2	1,578	26,297	69,479	,990	16,497	54,469
3	,697	11,611	81,090			
4	,582	9,698	90,788			
5	,295	4,916	95,704			
6	,258	4,296	100,000			

Extraktionsmethode: Hauptachsen-Faktorenanalyse.

Faktorenmatrix[a]

	Faktor 1	2
t6_paragraph_comprehension	,843	-,104
t7_sentence	,843	-,233
t9_word_meaning	,817	-,137
t14_word_recognition	,331	,590
t15_number_recognition	,152	,587
t17_object_number	,241	,462

Extraktionsmethode: Hauptachsenfaktorenanalyse.

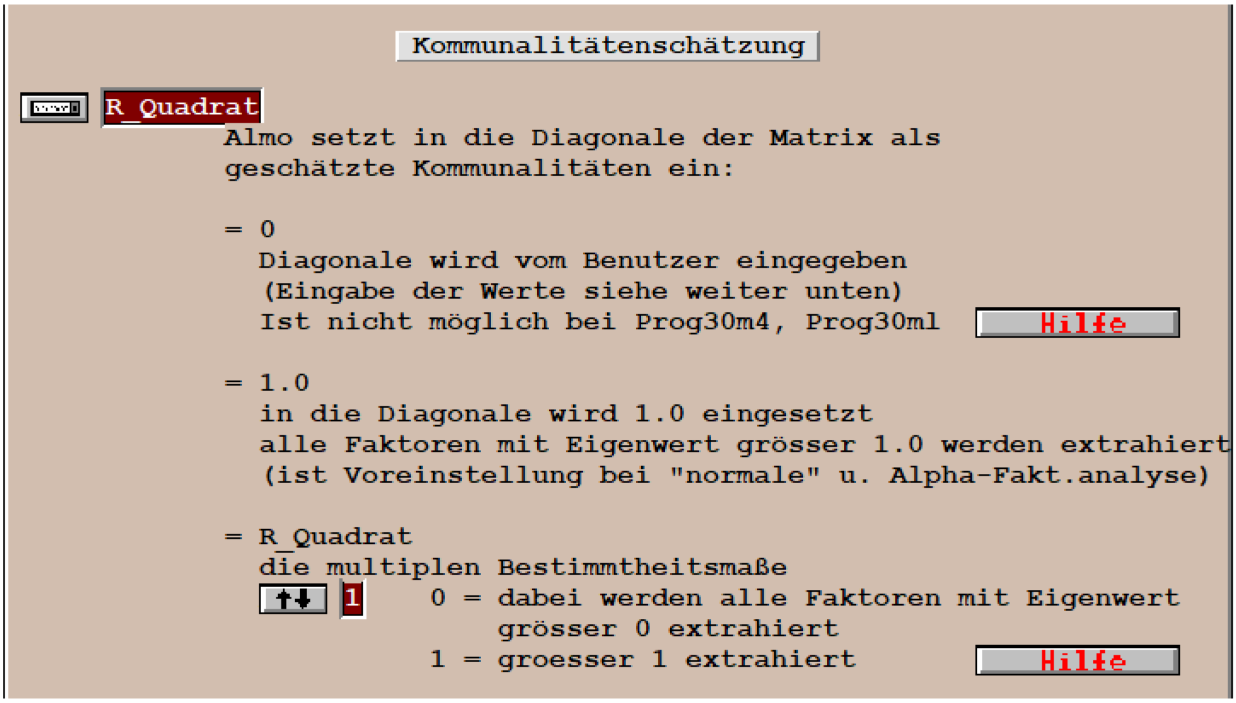
a. 2 Faktoren extrahiert. Es werden 10 Iterationen benötigt.

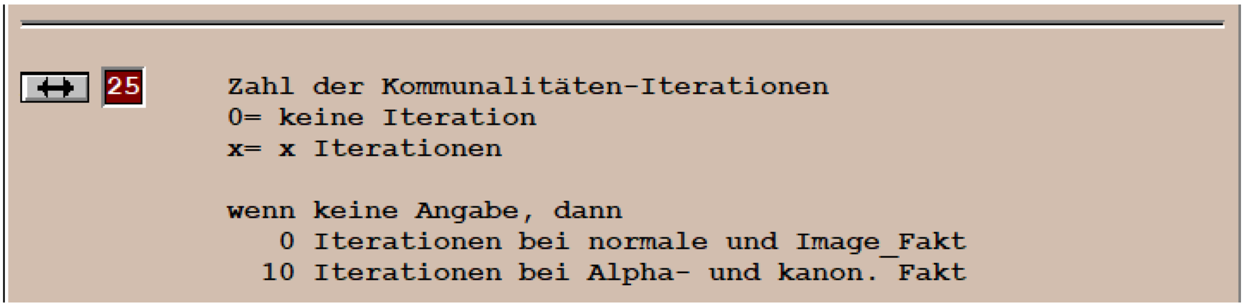
Am Unterrand der Tabelle 2 steht zwar "Exraktionsmethode: Haupt*achsen*-Faktorenanalyse". Diese Anmerkung bezieht sich auf den hinteren Tabellenteil. Im vorderen Teil werden alle, d.h. 6 Eigenwerte der Korrelationsmatrix mit 1.0 in der Diagonalen ausgegeben.

Die 2. Vorgehensweise, die SPSS dem Benutzer bei der Hauptkomponenten-Methode anbietet ist
"Feste Anzahl von Faktoren"
Der Benutzer gibt die Zahl der Faktoren an, die er extrahieren möchte. Da es aber darum geht, die Zahl der *inhaltlich signifikanten* d.h. "hinsichlich des Forschungsgegenstandes inhaltlich interpretierbaren" Faktoren zu finden, ist das keine einfache Entscheidung. In aller Regel wird man eine oder sogar mehrere weitere Analysen rechnen müssen, um schlussendlich, mit einer *plausiblen inhaltlich signifikanten* Faktorenzahl zu rechnen.

A2.4 Almo: Hauptachsenanalyse

In **Almo** kann die beschriebene SPSS-Vorgehensweise *"Basierend auf dem Eigenwert"* folgendermaßen realisiert werden. Die Optionsbox "Kommunalitätenschätzung" wird geöffnet. Wir bilden sie hier nur ausschnittsweise ab





Wird im 1. Eingabefeld als Kommunalitätenschätzung R_QUDRAT (multiples Bestimmtheitsmaß) eingesetzt, dann werden dem Benutzer im 2. kleinen Eingabefeld 2 weitere Möglichkeiten angeboten

```
0 = es werden alle Faktoren mit Eigenwert größer 0 extrahiert
    Dabei werden die Eigenwerte berechnet aus der Koorelationsmatrix
    mit den multiplen Bestimmtheitsmassen in der Diagonalen.
    Das ist das "Guttman-Kriterium" für die Kommunalitätenschätzung,
    das nur bei kleineren Korrelationsmatrizen erfolgreich eingesetzt werden
    kann. Bei größeren Korrelationsmatrizen erzeugt es zu viele Faktoren.

1 = es werden alle Faktoren mit Eigenwert größer 1 extrahiert
    Dabei werden die Eigenwerte berechnet aus der Koorelationsmatrix
    mit 1.0 in der Diagonalen. Danach werden die multiplen Bestimmtheits-
    maße in die Diagonale der Korrelationsmatrix eingesetzt und die
    Faktorenanalyse gerechnet.
    Das entspricht der in SPSS angebotenen Vorgehensweise, wenn dort
    die Methode "Hauptachsenanalyse" gewählt wurde
    und "basierend auf Eigenwert größer 1" eingesetzt wurde.
```

Almo liefert folgendes Ergebnis (gekürzt und etwas umgestellt):

```
Eigenwerte groesser 0 aus Korrelationsmatrix mit 1.0 in der Diagonalen
    2.59090      1.57785      0.69667      0.58188      0.29496      0.25775

Eigenwerte > 0 =6    Eigenwerte > 1 =2

2 Eigenwerte der Korrelationsmatrix (mit Diagonale 1.0) sind groesser 1.0
```

Entsprechend versucht Almo 2 Faktoren fuer die nachfolgende Faktorenanalyse zu extrahieren

Eingesetzte Werte fuer Kommunalitaeten (Multiple Bestimmtheitsmasse)

t6_parag	V14	0.6111
t7_sente	V15	0.6252
t9_word_	V17	0.5887
t14_word	V22	0.2453
t15_numb	V23	0.1990
t17_obje	V25	0.1665

Ergebnisse nach 9 Kommunalitaeten-Iterationen

Eigenwerte (Varianz je Faktor)
----

alle Eigenwerte > 0
2.2783 0.9898

fuer Analyse verwendete Eigenwerte
2.2783 0.9898

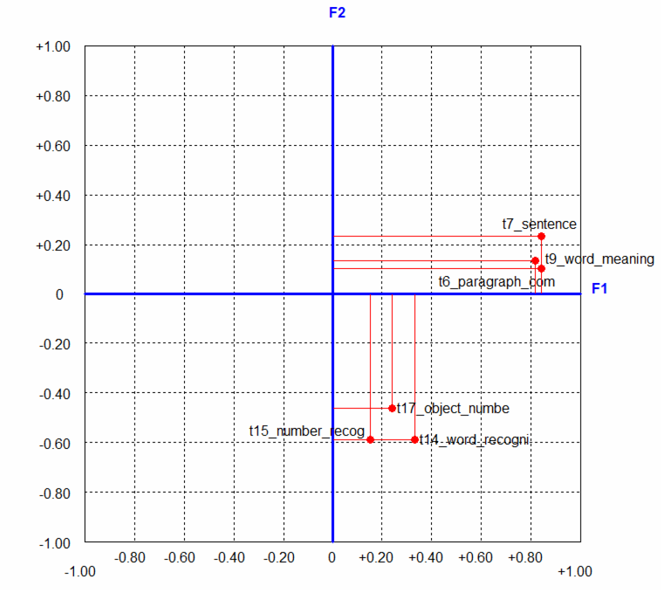
Prozent der Varianz je Faktor
37.9715 16.4974

Zu erklaerende Gesamtvarianz= 6.0000
Durch 2 Faktoren erklaerte Varianz= 3.2681
Prozentsatz der erklaerten Varianz= 54.4689

Matrix der Faktorladungen

		Faktor 1	Faktor 2
t6_parag	V14	0.8430	0.1039
t7_sente	V15	0.8428	0.2327
t9_word_	V17	0.8166	0.1367
t14_word	V22	0.3309	-0.5895
t15_numb	V23	0.1517	-0.5873
t17_obje	V25	0.2406	-0.4623

Grafisch dargestellt:



reproduzierte Kommunalitaeten je Variable

t6_parag	0.7215
t7_sente	0.7645
t9_word_	0.6856
t14_word	0.4571
t15_numb	0.3680
t17_obje	0.2716

Zahl der Kommunalitaeten-Iterationen: 9
Letzte Iterationsdifferenz je Variable

t6_parag	V14	0.0001
t7_sente	V15	0.0004
t9_word_	V17	0.0002
t14_word	V22	0.0008
t15_numb	V23	0.0003
t17_obje	V25	0.0003

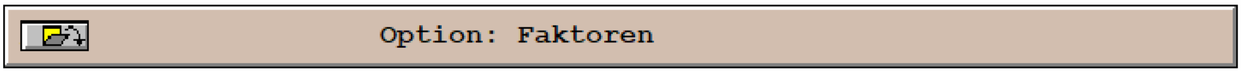
Konvergenz der Kommunalitaeten-Iterationen

Kommunalitaeten-Differenz	t6_para	t7_sent	t9_word	t14_wor	t15_num	t17_obj
von Iteration -1 zu 0	0.077	0.082	0.072	0.117	0.107	0.092
von Iteration 0 zu 1	0.025	0.031	0.021	0.047	0.039	0.021
von Iteration 1 zu 2	0.008	0.013	0.005	0.021	0.015	0.002
von Iteration 2 zu 3	0.002	0.006	0.001	0.011	0.006	0.003
von Iteration 3 zu 4	0.000	0.003	0.001	0.006	0.002	0.003
von Iteration 4 zu 5	0.000	0.002	0.001	0.004	0.001	0.002
von Iteration 5 zu 6	0.000	0.001	0.001	0.003	0.000	0.001
von Iteration 6 zu 7	0.000	0.001	0.000	0.002	0.000	0.001
von Iteration 7 zu 8	0.000	0.001	0.000	0.001	0.000	0.000
von Iteration 8 zu 9	0.000	0.000	0.000	0.001	0.000	0.000
konvergiert?	ja	ja	ja	ja	ja	nein

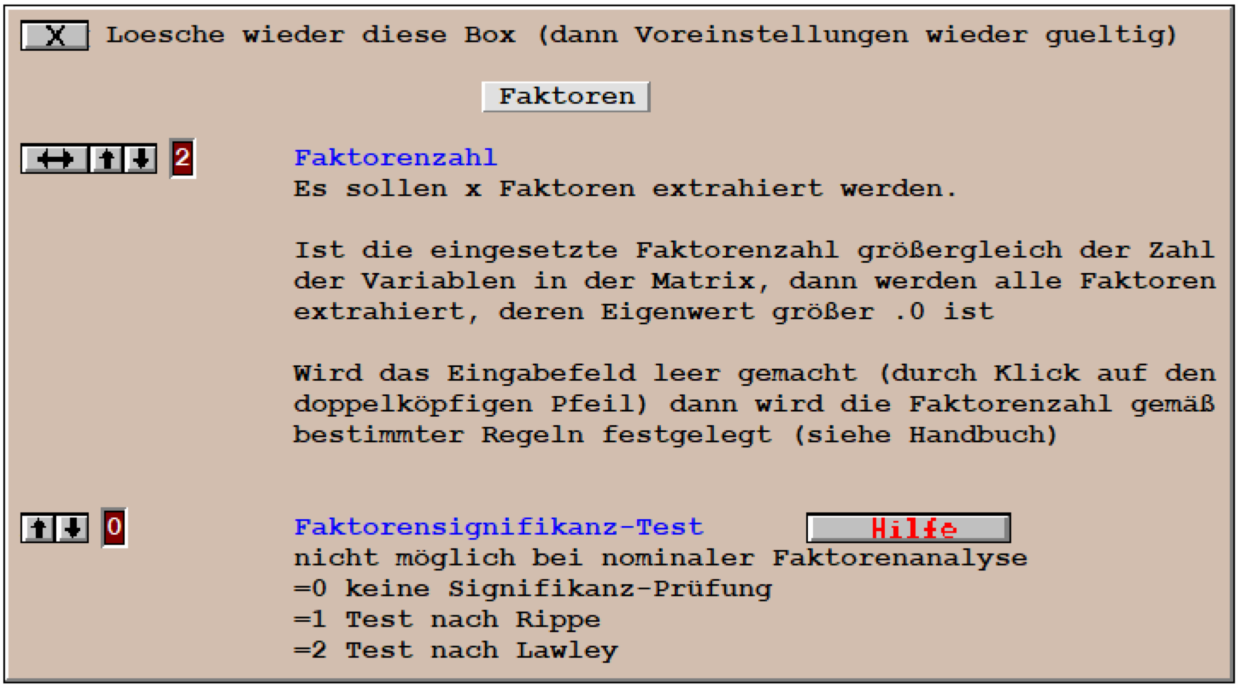
Die letzte Variable t17 konvergiert nicht exakt. Vom 2. Iterationswert 0.002 steigt die Kommunalitäten-Differenz wieder auf 0.003 an. Das ist jedoch so minimal, dass es negiert werden kann. Wenn die beiden Faktoren nicht so deutlich voneinder getrennt sind, wie in unserem Beispiel, dann wird man mit einer eher stark "gestörten" Konvergenz rechnen müssen.

In **Almo** (Prog30m2.Msk) wird die SPSS-Vorghensweise *"Feste Anzahl von Faktoren"* mit folgenden entsprechenden Einstellungen gerechnet.

Die Optionsbox "Faktoren"



muss geöffnet werden.



Bei gleicher vorgegebener Faktorenzahl sind die unrotierten Faktorladungsmatrizen aus Almo und SPSS exakt identisch. Wird rotiert dann muss, um vergleichen zu können, in SPSS auf Oblimin 0 und in Almo auf "nur Quartimin-Lösung" eingestellt werden. Die beiden Rotations-Lösungen sind minimal verschieden. Dem Begriff "Mustermatrix" in SPSS entspricht der Begriff "Ladungsmatrix" in Almo.

A2.5 Eine spezielle Konstellation

Wir rechnen mit SPSS eine Hauptachsenanalyse mit
"basierend auf Eigenwert größer als 0" und
ohne Kommunalitäten-Iterationen.
Im untersten Kästchen müssen wir dann 1 anstelle 25 einsetzen.
Äquivalent wäre auch eine Hauptachsenanalyse möglich mit folgender Eingabe
Feste Anzahl von Faktoren = 5 und
ohne Kommunalitäten-Iterationen.

Die Kommunalitäten-Iterationen setzen wir nur deswegen auf 0, damit die SPSS- und Almo-Ergebnisse exakt verglichen werden können.
Bei der ersten Variante wird 0 von SPSS interpretiert als "alle Faktoren". Für SPSS sind das 6-1=5. Das Ergebnis aus beiden Varianten ist identisch

Almo und SPSS geben dann aus:

<u>Das Almo-Ergebnis</u>

Almo gibt zuerst die Eigenwerte aus der Korrelationsmatrix mit 1.0 in der Diagonalen aus

```
Eigenwerte groesser 0 aus Korrelationsmatrix mit 1.0 in der Diagonalen
    2.59090      1.57785      0.69667      0.58188      0.29496      0.25775

Eigenwerte > 0 =6    Eigenwerte > 1 =2

2 Eigenwerte der Korrelationsmatrix (mit Diagonale 1.0) sind groesser 1.0
Entsprechend versucht Almo 2 Faktoren fuer die nachfolgende
Faktorenanalyse zu extrahieren

Eingesetzte Werte fuer Kommunalitaeten
```

(Multiple Bestimmtheitsmasse)

t6_parag	0.6111
t7_sente	0.6252
t9_word_	0.5887
t14_word	0.2453
t15_numb	0.1990
t17_obje	0.1665

Dann werden die Eigenwerte (bzw. Varianzen) aus der Korrelationsmatrix mit R_QUADRAT in der Diagonalen mitgeteilt

```
Eigenwerte (Varianz je Faktor)
--------------------------------
Eigenwerte > 0
    2.1579      0.8248     -0.2005     Beachte: Schon der 3.Eigenwert ist negativ !!

Prozent der Varianz je Faktor
   35.9656     13.7469

Zu erklaerende Gesamtvarianz=          6.0000
Durch 2 Faktoren erklaerte Varianz=    2.9827
Prozentsatz der erklaerten Varianz=   49.7125
```

Matrix der Faktorladungen

	Faktor 1	Faktor 2
t6_parag	0.8246	0.0888
t7_sente	0.8149	0.2068
t9_word_	0.8039	0.1214
t14_word	0.3060	-0.5182
t15_numb	0.1409	-0.5353
t17_obje	0.2328	-0.4520

Das SPSS-Ergebnis

Tabelle 1

Kommunalitäten

	Anfänglich	Extraktion
t6_paragraph_comprehension	,611	,764
t7_sentence	,625	,778
t9_word_meaning	,589	,726
t14_word_recognition	,245	,374
t15_number_recognition	,199	,344
t17_object_number	,167	,343

Extraktionsmethode: Hauptachsen-Faktorenanalyse.

Tabelle 2

Erklärte Gesamtvarianz

	Anfängliche Eigenwerte			Summen von quadrierten Faktorladungen für Extraktion		
Faktor	Gesamt	% der Varianz	Kumulierte %	Gesamt	% der Varianz	Kumulierte %
1	2,591	43,182	43,182	2,158	35,966	35,966
2	1,578	26,297	69,479	,825	13,747	49,712
3	,697	11,611	81,090	,094	1,562	51,275
4	,582	9,698	90,788	,118	1,964	53,239
5	,295	4,916	95,704	,135	2,248	55,487
6	,258	4,296	100,000			

Extraktionsmethode: Hauptachsen-Faktorenanalyse.

Im hinteren Teil der 2. Tabelle werden die Eigenwerte aus der eigentlichen Analyse mit den "anfänglichen" Kommunalitäten, den R_Quadrat ausgegeben. SPSS verwendet anstelle des Begriffs "Eigenwert" die Bezeichnung "Summe der quadrierten Faktorladungen" bzw. "Varianz". Merkwürdigerweise steigen ab dem 3. Eigenwert die Werte wieder an - was nicht möglich ist. Der SPSS-Kalkül erkennt nicht, dass schon der 3. Eigenwert negativ ist. Die ersten beiden Faktoren stimmen mit denen von Almo exakt überein. Die nachfolgenden Faktoren 3,4,5 sind falsch. Sie sind nicht existent.

Faktorenmatrix[a]

	Faktor				
	1	2	3	4	5
t6_paragraph_comprehension	,825	-,089	-,168	-,125	-,178
t7_sentence	,815	-,207	-,040	,226	,137
t9_word_meaning	,804	-,121	,211	-,134	,054
t14_word_recognition	,306	,518	-,071	,037	,076
t15_number_recognition	,141	,535	-,025	-,115	,154
t17_object_number	,233	,452	,117	,137	-,228

Extraktionsmethode: Hauptachsenfaktorenanalyse.

a. Es wurde versucht, 5 Faktoren zu extrahieren. Es werden mehr als 1 Iterationen benötigt. (Konvergenz=,176). Die Extraktion wurde abgebrochen.

A2.6 Weitere faktoranalytische Methoden

SPSS verfügt noch über diese Methoden:
Nicht gewichtete Kleinste Quadrate (ULS-Faktorenanalyse)
Verallgemeinerte Kleinste Quadrate (GLS-Faktorenanalyse)
Maximum Likelihood (ML-Faktorenanalyse)
Alpha-Faktorenanalyse
Image-Faktorenanalyse

Almo verfügt über diese:
Kanonische Faktorenanalyse
Alpha-Faktorenanalyse
Image-Faktorenanalyse

Es können also nur die Alpha- und die Image-Faktorenanalyse verglichen werden. Wir werden allerdings auch die Maximum-Likelihood-Faktorenanalyse aus SPSS und die

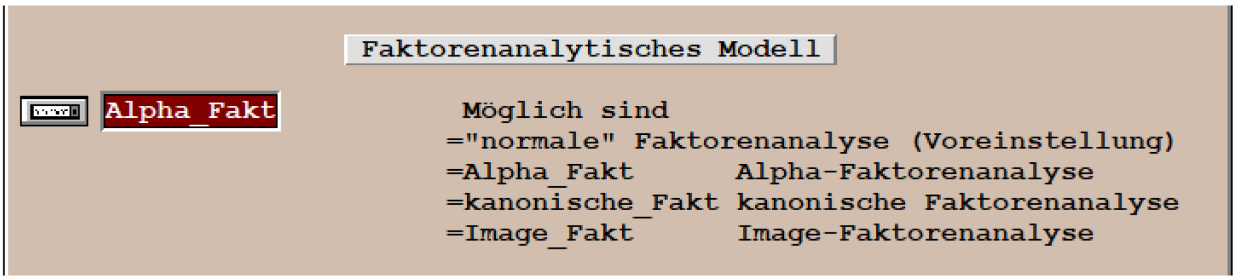
Kanonische Faktorenanalyse aus Almo miteinander vergleichen, da letztere als eine Variante der ML-Faktorenanalyse betrachtet werden kann. Da die GLS-Faktorenanalyse in SPSS sehr ähnliche Ergebnisse wie die ML-Faktorenanalyse erbringt ist damit auch indirekt ein Vergleich zur kanonischen Faktorenanalyse in Almo hergestellbar.

A2.6.1 Alpha-Faktorenanalyse

Der Kalkül der Alpha-Faktorenanalyse ist sehr komplex. In Anhang A1.2 haben wir ihn dargestellt.

SPSS. Als anfängliche Kommunalitätenschätzung werden die multiplen Bestimmtheitsmaße eingesetzt, dann wird iteriert. SPSS ermittelt die Zahl z der Eigenwerte der *Korrelatonsmatrix*, die größer als 1 (bzw. ein anderer Grenzwert) sind. Für diese z Faktoren wird dann der Kommunalitäten-Iterations-Kalkül für die jeweils neu errechneten *Alphamatrizen* durchgeführt.

Almo. In Almo muss die Optionsbox "Faktorenanalytisches Modell" geöffnet werden. Wir zeigen einen Ausschnitt aus der Optionsbox



Almo rechnet die Alpha-Faktorenanalyse mit 1.0 als anfänglichen Kommunalitätenschätzungen.

Um wie SPSS mit den multiplen Bestimmtheitsmaßen zu rechnen, müsste in Almo zusätzlich die Optionsbox "Kommunalitätenschätzung" geöffnet werden und R_QUADRAT eingesetzt werden. *Wir raten davon ab*, da je nach Datenkonstellation eine nicht positiv semidefinite Matrix entstehen kann.

Almo bildet die 1. Alphamatrix und berechnet deren Eigenwerte. Für die Faktorenzahl mit Eigenwerten grösser 1 wird dann die Kommunalitäten-Iteration für die jeweils neu errechneten *Alphamatrizen* durchgeführt und dann abschließend die Faktorladungsmatrix errechnet. Die 1. Alphamatrix ist (bei 1.0 als anfänglichen Kommunalitäten) identisch mit der Korrelationsmatrix mit 1.0 in der Diagonalen.

Die Ergebnisse von SPSS und Almo stimmen exakt überein, differieren vielleicht an der 3. Kommastellen - je danach ob gleich viel oder ungleich viele Iterationszyklen gerechnet wurden und welche Schwellenwerte verwendet wurden.

<u>Ergebnisse aus SPSS</u>

Kommunalitäten

	Anfänglich	Extraktion
t6_paragraph_comprehension	,611	,720
t7_sentence	,625	,768
t9_word_meaning	,589	,685
t14_word_recognition	,245	,462
t15_number_recognition	,199	,362
t17_object_number	,167	,272

Extraktionsmethode: Alpha-Faktorisierung.

Erklärte Gesamtvarianz

	Anfängliche Eigenwerte			Summen von quadrierten Faktorladungen für Extraktion		
Faktor	Gesamt	% der Varianz	Kumulierte %	Gesamt	% der Varianz	Kumulierte %
1	2,591	43,182	43,182	2,018	33,642	33,642
2	1,578	26,297	69,479	1,250	20,829	54,471
3	,697	11,611	81,090			
4	,582	9,698	90,788			
5	,295	4,916	95,704			
6	,258	4,296	100,000			

Extraktionsmethode: Alpha-Faktorisierung.

Faktorenmatrix[a]

	Faktor	
	1	2
t6_paragraph_comprehension	,705	,473
t7_sentence	,648	,590
t9_word_meaning	,670	,485
t14_word_recognition	,562	-,381
t15_number_recognition	,397	-,452
t17_object_number	,423	-,305

Extraktionsmethode: Alpha-Faktorisierung.

a. 2 Faktoren extrahiert. Es werden 7 Iterationen benötigt.

Ergebnisse aus Almo

Übereinstummung mit SPSS

```
Eigenwerte groesser 0 aus Korrelationsmatrix mit 1.0 in der Diagonalen
    2.59090      1.57785      0.69667      0.58188      0.29496      0.25775
Eigenwerte > 0 =6    Eigenwerte > 1 =2
```

Alphamatrix nach 10 Kommunalitaeten-Iterationen

wird bei SPSS nicht ausgegeben

	t6_para	t7_sent	t9_word	t14_wor	t15_num	t17_obj
t6_parag	1.000	0.986	1.004	0.387	0.136	0.329
t7_sente	0.986	1.000	0.993	0.258	-0.036	0.202
t9_word_	1.004	0.993	1.000	0.307	0.104	0.337
t14_word	0.387	0.258	0.307	1.000	0.968	1.003
t15_numb	0.136	-0.036	0.104	0.968	1.000	0.969
t17_obje	0.329	0.202	0.337	1.003	0.969	1.000

Eigenwert der "Alphamatrix" je Faktor *wird bei SPSS nicht ausgegeben*

```
    3.671      2.329     -0.058
```

Varianz je Faktor

```
    2.019      1.250
```

Übereinstimmung mit SPSS

Prozent der Varianz

```
   33.658     20.827
```

Übereinstimmung mit SPSS

```
Zu erklaerende Gesamtvarianz=        6.000
Durch 2 Faktoren erklaerte Varianz=  3.269
Prozentsatz der erklaerten Varianz= 54.485
```

Übereinstimmung mit SPSS

Alpha-Koeffizient je Faktor *wird bei SPSS nicht ausgegeben*

```
    0.728      0.571
```

Matrix der Faktorladungen

	Faktor 1	Faktor 2
t6_parag	0.705	0.472
t7_sente	0.649	0.589
t9_word_	0.671	0.484
t14_word	0.562	-0.381
t15_numb	0.398	-0.454
t17_obje	0.422	-0.305

Übereinstimmung mit SPSS

reproduzierte Kommunalitaeten je Variable

t6_parag	V14	0.719
t7_sente	V15	0.769
t9_word_	V17	0.684
t14_word	V22	0.461
t15_numb	V23	0.365
t17_obje	V25	0.271

Übereinstimmung mit SPSS
wird in Tabelle "Kommunalitäten
in Spalte "Extraktion" ausgegeben

Kommunalitaeten-Iterationen: 10
Letzte Iterationsdifferenz (je Variable)

t6_parag	V14	0.000
t7_sente	V15	0.000
t9_word_	V17	0.000
t14_word	V22	0.001
t15_numb	V23	0.000
t17_obje	V25	0.000

wird bei SPSS nicht ausgegeben

wird bei SPSS nicht ausgegeben

Konvergenz der Kommunalitaeten-Iterationen:

Kommunalitaeten-Differenz	t6_para	t7_sent	t9_word	t14_wor	t15_num	t17_obj
von Iteration -1 zu 0	0.189	0.170	0.202	0.387	0.396	0.487
von Iteration 0 zu 1	0.062	0.052	0.069	0.126	0.146	0.152
von Iteration 1 zu 2	0.020	0.014	0.025	0.036	0.054	0.053
von Iteration 2 zu 3	0.006	0.002	0.010	0.006	0.021	0.020
von Iteration 3 zu 4	0.002	0.001	0.004	0.002	0.009	0.009
von Iteration 4 zu 5	0.001	0.002	0.002	0.004	0.004	0.004
von Iteration 5 zu 6	0.000	0.001	0.001	0.003	0.002	0.002
von Iteration 6 zu 7	0.000	0.001	0.001	0.003	0.001	0.001
von Iteration 7 zu 8	0.000	0.001	0.000	0.002	0.001	0.001
von Iteration 8 zu 9	0.000	0.000	0.000	0.001	0.001	0.000
von Iteration 9 zu 10	0.000	0.000	0.000	0.001	0.000	0.000
konvergiert?	ja	nein	ja	nein	ja	ja
mit Toleranz von 0.009	ja	ja	ja	ja	ja	ja

A2.6.2 Image-Faktorenanalyse

Die Ergebnisse von SPSS und Almo sind identisch. In Almo muss die Optionsbox "Faktorenanalytisches Modell" geöffnet werden. Wir zeigen einen Ausschnitt aus der Optionsbox

```
                    Faktorenanalytisches Modell

 [....]  Image_Fakt          Möglich sind
                            ="normale" Faktorenanalyse (Voreinstellung)
                            =Alpha_Fakt       Alpha-Faktorenanalyse
                            =kanonische_Fakt  kanonische Faktorenanalyse
                            =Image_Fakt       Image-Faktorenanalyse
```

SPSS gibt standardmäßig zusätzlich noch die "Kovarianzmatrix" aus. In Almo wird sie unter dem Namen "unskalierte Imagematrix" ausgegeben.

Almo rechnet immer ohne Kommunalitäten-Iteration. Auch SPSS rechnet ohne Iteration. Dies auch wenn die Iterationszahl auf 25 gesetzt ist.

Wird nicht die Korrelationsmatrix sondern die *Kovarianzanalyse* der Image-Faktorenanalyse unterzogen, dann stellt sich die Frage, wie viele Faktoren sollen extrahiert werden?

In **Almo** muss dazu die Optionsbox „Zu faktorisierende Matrix“ geöffnet werden und auf „Kovarianz“ umgestellt werden.

```
 [ X ] Loesche wieder diese Box (dann Voreinstellungen wieder gueltig)

                    zu faktorisierende Matrix

 [....]  Kovarianz
                                          Faktorisiert wird
                                          -----------------
                    =Korrelation          Korrelationsmatrix
                    =Quadratsumme         Quadratsummenmatrix
                    =Kovarianz            Kovarianzmatrix
                    =Quasikorrelation     s. Handbuch, P20.8.5.1
```

Almo überführt die Kovarianzmatrix vorübergehend in die Korrelationsmatrix mit 1.0 in der Diagonalen und extrahiert – wie gewohnt nach dem Kaiser-Kriterium – alle Faktoren mit Eigenwert grösser 1. Wir zeige zuerst das **Almo-Ergebnis** wobei die Ergebnisse für die Image-Faktorenanalyse für die Kovarianzmatrix und die Korrelationsmatrix nebeneinander gestellt werden

```
                    Ergebnis aus Image-Fakorenanalyse für
            Kovarianzmatrix                  |          Korrelationsmatrix
-----------------------------------------    |---------------------------------------
Summe der Diagonale der Kovarianzmatrix      |Summe der Diagonale der Korrelationsmatrix
mit Varianzen in der Diagonalen = 313.991    |mit 1.0 in Diagonale = 6.0
                                             |
Eigenwerte groesser 0                        |Eigenwerte groesser 0
  156.43481                                  |  2.59090
  78.26843                                   |  1.57785
  45.33328                                   |  0.69667
  19.43871                                   |  0.58188
  10.42074                                   |  0.29496
```

```
  4.09479                                    |  0.25775
                                             |
Summe der Eigenwert groesser 0 = 313.991     |Summe der Eigenwert groesser 0 = 6.0
                                             |
Um die zu extrahierende Zahl der Faktoren    |
festzulegen wird die Kovarianzmatrix         |   ----
temporaer in die Korrelationsmatrix          |
umgerechnet                                  |
                                             |
2 Eigenwerte der Korrelationsmatrix          |2 Eigenwerte der Korrelationsmatrix
sind groesser 1.0. Entsprechend versucht     |(mit Diagonale 1.0) sind groesser 1.0
Almo 2 Faktoren fuer nachfolgende            |Entsprechend versucht Almo 2 Faktoren
Faktorenanalyse der Kovarianzmatrix          |fuer die nachfolgende Faktorenanalyse
zu extrahieren                               |zu extrahieren
                                             |
                     alle Eigenwerte aus (skalierter) Image-Matrix > 0
                     für Kovarianz- und Korrelations-Matrix gleich
                                        4.5083
                                        0.5611
                                        0.1716
                                        0.1203
                                        0.0579
                                        0.0246
                                             |
                 Kommunalitäten              |                 Kommunalitäten
             eingesetzt |reproduziert        |             eingesetzt |reproduziert
    +----------+----------+--------+         |     +----------+----------+--------+
    |t6_parag  |   7.453  |  7.059 |         |     |t6_parag  |   0.611  |  0.579 |
    |t7_sente  |  16.658  | 15.647 |         |     |t7_sente  |   0.625  |  0.587 |
    |t9_word_  |  34.626  | 32.899 |         |     |t9_word_  |   0.589  |  0.559 |
    |t14_word  |  32.478  | 27.372 |         |     |t14_word  |   0.245  |  0.207 |
    |t15_numb  |  11.889  |  9.770 |         |     |t15_numb  |   0.199  |  0.164 |
    |t17_obje  |   4.025  |  3.556 |         |     |t17_obje  |   0.167  |  0.147 |
    +----------+----------+--------+         |     +----------+----------+--------+
                                             |
Varianz je Faktor                            |Varianz je Faktor
   63.561                                    |    1.806
   32.742                                    |    0.436
                                             |
Prozent der Varianz je Faktor                |Prozent der Varianz je Faktor
   20.243                                    |   30.104
   10.428                                    |   7.275
                                             |
Zu erklaerende Gesamtvarianz=     313.991    |Zu erklaerende Gesamtvarianz=        6.000
Durch 2 Faktoren erklaerte Varianz=96.303    |Durch 2 Faktoren erklaerte Varianz=  2.243
Prozentsatz der erklaerten Varianz=30.671    |Prozentsatz der erklaerten Varianz= 37.379
                                             |
Matrix der Faktorladungen                    |Matrix der Faktorladungen
            +--------------------+           |            +--------------------+
            | Faktor 1 Faktor 2  |           |            | Faktor 1 Faktor 2  |
  +---------+--------------------+           |  +---------+--------------------+
  |t6_parag |    2.657    0.019  |           |  |t6_parag |    0.761    0.006  |
  |t7_sente |    3.934    0.416  |           |  |t7_sente |    0.762    0.081  |
  |t9_word_ |    5.731    0.221  |           |  |t9_word_ |    0.747    0.029  |
  |t14_word |    2.676   -4.496  |           |  |t14_word |    0.233   -0.391  |
  |t15_numb |    0.570   -3.073  |           |  |t15_numb |    0.074   -0.398  |
  |t17_obje |    0.834   -1.692  |           |  |t17_obje |    0.170   -0.344  |
  +---------+--------------------+           |  +---------+--------------------+
```

Das graphische Abbild der beiden Faktorladungsmatrizen ist deckungsgleich

Das SPSS-Ergebnis

Bei SPSS wird nicht ersichtlich, wie die Zahl der zu extrahierenden Faktoren für die Kovarianzanalyse bestimmt wird. SPSS extrahiert 4 Faktoren. Da die Imagematrix positiv semidefinit ist, könnten aus ihr soviele Faktoren extrahiert werden, wie Variable vorhanden sind. In unserem Beispiel sind das 6. SPSS ist nur bereit 6-1=5 Faktoren zu extrahieren.

Tabelle 1

Kommunalitäten

	Unbearbeitet		Neu skaliert	
	Anfänglich	Extraktion	Anfänglich	Extraktion
t6_paragraph_comprehension	7,453	7,073	,611	,580
t7_sentence	16,658	15,674	,625	,588
t9_word_meaning	34,626	33,301	,589	,566
t14_word_recognition	32,478	29,368	,245	,222
t15_number_recognition	11,889	10,878	,199	,182
t17_object_number	4,025	3,898	,167	,161

Extraktionsmethode: Image-Faktorisierung.

Der vordere Teil der Tabelle, überschrieben mit „Unbearbeitet“ bringt das Ergebnis für die Kovarianzmatrix, der hintere Teil, überschrieben mit „Neu skaliert“ das Ergebnis für die Korrelationsmatrix.
Die jeweils "anfänglichen Kommunalitäten" von SPSS und Almo stimmen überein. Die mit "Extraktion" überschriebenen weichen voneinder ab, da SPSS mit 4 Faktoren rechnet.

Tabelle 2

Erklärte Gesamtvarianz

		Anfängliche Eigenwerte[a]			Summen von quadrierten Faktorladungen für Extraktion		
	Faktor	Gesamt	% der Varianz	Kumulierte %	Gesamt	% der Varianz	Kumulierte %
Unbearbeitet	1	156,435	49,821	49,821	63,561	59,332	59,332
	2	78,268	24,927	74,748	32,742	30,563	89,895
	3	45,333	14,438	89,186	,741	,692	90,587
	4	19,439	6,191	95,377	3,148	2,938	93,525
	5	10,421	3,319	98,696			
	6	4,095	1,304	100,000			
Neu skaliert	1	156,435	49,821	49,821	1,806	30,104	30,104
	2	78,268	24,927	74,748	,436	7,275	37,379
	3	45,333	14,438	89,186	,020	,327	37,706
	4	19,439	6,191	95,377	,037	,619	38,325
	5	10,421	3,319	98,696			
	6	4,095	1,304	100,000			

Extraktionsmethode: Image-Faktorisierung.

a. Gleiche Start-Eigenwerte für die unbearbeitete und neuskalierte Lösung bei der Analyse der Kovarianzmatrix .

In der vorderen Hälfte der Tabelle, die mit "Anfängliche Eigenwerte" überschrieben ist, werden die Eigenwerte der Kovarianzmatrix (mit den Varianzen der Variablen in der

Diagonalen) ausgegeben. Almo liefert dasselbe Ergebnis. In der hinteren Hälfte werden - in der Terminologie von Almo - die Eigenwerte der (skalierten) Imagematrix in der Spalte "Gesamt" ausgegeben. Sie werden bei SPSS als "Summe der quadrierten Faktorladungen" bezeichnet. Die Werte sind 63.561, 32.742. Sie stimmen exakt mit Almo überein. Danach erfolgt ein eklatanter Absturz auf .741 - was ein indirekter Hinweis darauf ist, dass nur 2 Faktoren und nicht 4 relevant sind. Dann steigt der Wert auf 3.148 an. Das ist zwar überraschend aber durchaus möglich, da die Daten in unterschiedlichen Maßeinheiten gemessen sind und in der Kovarianzanalyse nicht standardisiert werden. Anders bei der Korrelationsmatrix. In obiger Tabelle 2 wird sie mit "Neu skaliert" bezeichnet. Hier sind die Werte 1,806, 0,436, 0.020, 0.037. Der Anstieg von 0.020 auf 0,037 ist nicht möglich und weist darauf hin, dass etwas nicht stimmt. Wird in Almo eine Analyse mit 4 Faktoren erzwungen, dann entstehen dort die Werte 1,806, 0,436, 0,076, 0,061.

Die Tabelle „Kovarianzmatrix für Image“, die wir hier nicht anzeigen, stimmen bei SPSS und Almo exakt überein. In Almo entspricht dem die "unskalierte Imagematrix".

Tabelle 3

Faktorenmatrix[a]

	Unbearbeitet Faktor				Neu skaliert Faktor			
	1	2	3	4	1	2	3	4
t6_paragraph_comprehension	2,657	-,019	-,061	-,102	,761	-,006	-,018	-,029
t7_sentence	3,934	-,416	-,001	-,165	,762	-,081	,000	-,032
t9_word_meaning	5,731	-,221	,113	,624	,747	-,029	,015	,081
t14_word_recognition	2,676	4,496	-,395	-1,356	,233	,391	-,034	-,118
t15_number_recognition	,570	3,073	-,476	,939	,074	,398	-,062	,121
t17_object_number	,834	1,692	,584	-,010	,170	,344	,119	-,002

Extraktionsmethode: Image-Faktorisierung.

a. 4 Faktoren extrahiert

Der vordere Teil der Tabelle, überschrieben mit „Unbearbeitet“ bringt das Ergebnis für die Kovarianzmatrix, der hintere Teil, überschrieben mit „Neu skaliert“ das Ergebnis für die Korrelationsmatrix.
Die jeweils ersten beiden Faktoren stimmen exakt mit denen von Almo überein.

A2.6.3 Kanonische (Almo) und Maximum-Likelihood Faktorenanalyse (SPSS)

Die in Almo enthaltene Kanonische Faktorenanalyse wird als äquivalent zur ML-Faktorenanalyse betrachtet (Arminger 1979, S.60). Harman (1976, S.219) schreibt: "The results ...are again maximum-likelihood estmates of facor loadings". Für unser Beispiel trifft das tatsächlich zu. Das muss aber nicht immer so sein. Wenn die Faktorenstruktur weniger eindeutig ist als in unserem Beispiel, dann kann die ML-Methode auch Ergebnisse erbringen, die unzulässig sind oder völlig anders sind als die Ergebnisse aller anderen Methoden. Die kanonische Faktorenanalyse ist dann immer ein zuverlässiges Verfahren. Wir rechnen das 2-faktorielle Beispiel mit den Daten von Holzinger/Swineford und ein zweites Beispiel mit allen Variablen t1 bis t24 mit 4 Faktoren.

Beispiel mit 2 Faktoren

Matrix der Faktorladungen

	Almo Faktor 1	Almo Faktor 2	SPSS F1	SPSS F2
t6_parag	0.850	0.010	,850	,010
t7_sente	0.864	-0.117	,865	-,117
t9_word_	0.829	-0.030	,828	-,029
t14_word	0.251	0.621	,251	,624
t15_numb	0.072	0.609	,071	,607
t17_obje	0.176	0.491	,176	,490

<u>Beispiel mit allen Variablen und 4 Faktoren</u>

SPSS extrahiert die Faktoren nach der Methode "Basierend auf Eigenwerten größer 1.0" - wobei die Eigenwerte aus der Korrelationsmatrix mit 1.0 in der Diagonalen gemeint sind. Diese Vorgehensweise entspricht nicht dem Modell der kanonischen Faktorenanalyse, führt aber in der Regel zu sinnvollen Faktorenzahlen.

Almo extrahiert die Eigenwerte aus der kanonischen Faktormatrix $U^{-1}(R-U^2)U^{-1}$ (siehe Abschnitt A1.1, Formel 3a). Die Zahl der Eigenwerte größer 0 ergibt dann die Zahl der zu extrahierenden Faktoren. Almo gelangt dabei zu 5 Faktoren. Um mit SPSS vergleichen zu können muss Almo gezwungen werden nur 4 Faktoren zu extrahieren. Das geschieht dadurch, das die Optionsbox "Faktoren geöffnet wird und 4 Faktoren angefordert werde.

Matrix der Faktorladungen

	Almo Faktor 1	Almo Faktor 2	Almo Faktor 3	Almo Faktor 4	SPSS F1	SPSS F2	SPSS F3	SPSS F4
t1_visua	0.526	0.232	0.315	-0.203	,525	,227	,318	-,204
t2_cubes	0.286	0.129	0.367	-0.208	,286	,123	,368	-,209
t3_paper	0.323	0.120	0.206	-0.235	,323	,117	,207	-,236
t4_lozen	0.336	0.374	0.371	-0.205	,336	,369	,375	-,207
t5_gener	0.767	-0.310	-0.121	-0.026	,767	-,309	-,125	-,025
t6_parag	0.785	-0.232	-0.043	0.081	,785	-,232	-,046	,081
t7_sente	0.791	-0.371	-0.096	0.027	,791	-,371	-,099	,027
t8_word_	0.725	-0.167	-0.011	0.025	,725	-,168	-,012	,025
t9_word_	0.810	-0.252	0.009	0.027	,810	-,252	,007	,026
t10_addi	0.295	0.444	-0.584	-0.064	,297	,455	-,584	-,063
t11_code	0.499	0.353	-0.284	0.002	,499	,356	-,276	,004
t12_coun	0.329	0.433	-0.313	-0.251	,330	,436	-,305	-,248
t13_stra	0.420	0.404	-0.103	-0.253	,420	,404	-,095	-,251
t14_word	0.299	0.264	0.113	0.522	,300	,262	,120	,523
t15_numb	0.168	0.373	0.182	0.379	,168	,371	,188	,378
t16_figu	0.424	0.370	0.254	0.195	,424	,366	,261	,194
t17_obje	0.256	0.413	-0.142	0.357	,257	,414	-,133	,357
t18_numb	0.307	0.335	0.011	0.222	,307	,334	,018	,222
t19_figu	0.390	0.189	0.090	0.176	,390	,187	,096	,176
t20_dedu	0.512	0.145	0.305	0.008	,512	,140	,307	,006
t21_nume	0.541	0.326	0.024	-0.131	,541	,326	,027	-,132
t22_prob	0.626	0.042	0.198	-0.066	,626	,039	,199	-,067
t23_seri	0.644	0.212	0.225	-0.125	,644	,209	,227	-,126
t24_wood	0.600	0.209	-0.114	0.047	,600	,211	-,111	,047

Die Unterschiede in beiden Beispielen sind vernachlässigbar gering.

A2.7 Rotation in Almo und SPSS

Rechtwinklige Rotation

Almo verwendet das Varimax-Verfahren. SPSS verwendet als Standarverfahren ebenfalls das Varimax-Verfahren, verfügt aber noch über das Equimax- und Quartimax-Verfahren. Das Varimax-Ergebnis aus Almo und SPSS ist exakt identisch

Schiefwinklige Rotation
Almo verwendet (a) das Quartimin-Verfahren und (b) das Gruppenrotations-Verfahren, das das Quartimin-Verfahren erweitert. In Abschnitt P30.2.3.2 wird es vorgestellt. Wenn sich die zu faktorisierenden Variablen relativ eindeutig in Gruppen trennen lassen, grafisch betrachtet, in "Punktewolken" separieren lassen, dann legt die Gruppenrotation die schiefwinkligen Achsen *varianzmaximierend* mitten durch die Punktewolken und liefert somit ein optimales Ergebnis.

SPSS verwendet als Standardverfahren (c) das Oblimin-Verfahren und bietet noch (d) das Promax-Verfahren an.
Die vier Verfahren a bis d liefern nur ähnliche aber nicht identische Ergebnisse.

Bei Oblimin in SPSS kann über den Eingabeparameter Delta (mit Werten von -9999 bis 0.8) der Winkel mit dem die schiefwinkligen Achsen aufeinander stehen (teilweise sehr drastisch) verändert werden. Einen besten oder sogar "wahren" Delta-Wert gibt es nicht. In SPSS ist Delta=0 voreingestellt.
Bei 2-faktoriellen Faktorladungsmatrizen sind die Rotationsergebnisse aus Quartimin und Oblimin mit Delta-Parameter 0 sehr ähnlich, wobei Oblimin etwas geringere Schiefe der Achsen erzeugt.

Literatur

Die Literatur zur Faktorenanalyse ist sehr umfangreich. Hier sollen deswegen nur einige wenige Arbeiten angegeben werden, die einen weiterführenden Charakter haben. Ein Klassiker, der (fast) den ganzen Bereich der Faktorenanalyse umfasst, ist das Buch von Harman.

Arminger, Gerhard: Faktorenanalyse, Teubner, Stuttgart 1979
Harman, H.H.: Modern Factor Analysis, University of Chicago Press, 3. Aufl. 1976
Holm, Kurt: Die Befragung, Band 3 "Faktorenanalyse", Francke Verlag, UTB 433, 1976
Überla, Kurt: Faktorenanalyse, Springer Verlag, Berlin Heidelberg New Yorg, 1977

Schlagwortverzeichnis